PLOS ONE

RESEARCH ARTICLE

Fine-Mapping of the 1p11.2 Breast Cancer Susceptibility Locus

Hisani N. Horne[1,2], Charles C. Chung[1], Han Zhang[1], Kai Yu[1], Ludmila Prokunina-Olsson[1], Kyriaki Michailidou[3], Manjeet K. Bolla[3], Qin Wang[3], Joe Dennis[3], John L. Hopper[4], Melissa C. Southey[5], Marjanka K. Schmidt[6], Annegien Broeks[6], Kenneth Muir[7,8], Artitaya Lophatananon[7], Peter A. Fasching[9,10], Matthias W. Beckmann[9], Olivia Fletcher[11,12], Nichola Johnson[11,12], Elinor J. Sawyer[13], Ian Tomlinson[14], Barbara Burwinkel[15,16], Frederik Marme[15,17], Pascal Guénel[18,19], Thérèse Truong[18,19], Stig E. Bojesen[20,21,22], Henrik Flyger[23], Javier Benitez[24,25], Anna González-Neira[24], Hoda Anton-Culver[26], Susan L. Neuhausen[27], Hermann Brenner[28,29,30], Volker Arndt[28], Alfons Meindl[31], Rita K. Schmutzler[32,33,34,35], Hiltrud Brauch[30,36,37], Ute Hamann[38], Heli Nevanlinna[39], Sofia Khan[39], Keitaro Matsuo[40], Hiroji Iwata[41], Thilo Dörk[42], Natalia V. Bogdanova[43], Annika Lindblom[44], Sara Margolin[45], Arto Mannermaa[46,47,48], Veli-Matti Kosma[46,47,48], Georgia Chenevix-Trench[49], kConFab/AOCS Investigators[50¶], Anna H. Wu[51], David ven den Berg[51], Ann Smeets[52], Hui Zhao[53,54], Jenny Chang-Claude[55,56], Anja Rudolph[55], Paolo Radice[57], Monica Barile[58], Fergus J. Couch[59], Celine Vachon[60], Graham G. Giles[4,61], Roger L. Milne[4,61], Christopher A. Haiman[51], Loic Le Marchand[62], Mark S. Goldberg[63,64], Soo H. Teo[65,66], Nur A. M. Taib[66], Vessela Kristensen[67,68,69], Anne-Lise Borresen-Dale[67,68], Wei Zheng[70], Martha Shrubsole[70], Robert Winqvist[71,72], Arja Jukkola-Vuorinen[73], Irene L. Andrulis[74,75], Julia A. Knight[76,77], Peter Devilee[78,79], Caroline Seynaeve[80], Montserrat García-Closas[1,81], Kamila Czene[82], Hatef Darabi[82], Antoinette Hollestelle[80], John W. M. Martens[80], Jingmei Li[82], Wei Lu[83], Xiao-Ou Shu[70], Angela Cox[84], Simon S. Cross[85], William Blot[70,86], Qiuyin Cai[70], Mitul Shah[87], Craig Luccarini[87], Caroline Baynes[87], Patricia Harrington[87], Daehee Kang[88,89,90], Ji-Yeob Choi[89,90], Mikael Hartman[91,92], Kee Seng Chia[91], Maria Kabisch[38], Diana Torres[38,93], Anna Jakubowska[94], Jan Lubinski[94], Suleeporn Sangrajrang[95], Paul Brennan[96], Susan Slager[60], Drakoulis Yannoukakos[97], Chen-Yang Shen[98,99], Ming-Feng Hou[100], Anthony Swerdlow[12,81], Nick Orr[11], Jacques Simard[101], Per Hall[82], Paul D. P. Pharoah[3,87], Douglas F. Easton[3,87☯], Stephen J. Chanock[1☯], Alison M. Dunning[87☯], Jonine D. Figueroa[1,102,103☯*]

1 Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, United States of America, **2** Food and Drug Administration, Silver Spring, MD, United States of America, **3** Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, **4** Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Melbourne, Australia, **5** Department of Pathology, The University of Melbourne, Melbourne, Australia, **6** Netherlands Cancer Institute, Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands, **7** Division of Health Sciences, Warwick Medical School, Warwick University, Coventry, UK, **8** Institute of Population Health, University of Manchester, Manchester, UK, **9** Department of Gynaecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, Erlangen, Germany, **10** David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA, United States of America, **11** Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK, **12** Division of Breast Cancer Research, The Institute of Cancer Research, London, UK, **13** Research Oncology, Guy's Hospital, King's College London, London, UK, **14** Wellcome Trust Centre for Human Genetics and Oxford Biomedical Research Centre, University of Oxford, Oxford, UK, **15** Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany, **16** Molecular Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany, **17** National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany, **18** Environmental Epidemiology of Cancer, Center for Research in Epidemiology and Population Health, INSERM, Villejuif, France, **19** University Paris-Sud, Villejuif, France, **20** Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark, **21** Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark, **22** Faculty of Health and Medical Sciences, University of

CrossMark
click for updates

OPEN ACCESS

Citation: Horne HN, Chung CC, Zhang H, Yu K, Prokunina-Olsson L, Michailidou K, et al. (2016) Fine-Mapping of the 1p11.2 Breast Cancer Susceptibility Locus. PLoS ONE 11(8): e0160316. doi:10.1371/journal.pone.0160316

Editor: Kelvin Yuen Kwong Chan, Hospital Authority, CHINA

Received: January 8, 2016

Accepted: July 18, 2016

Published: August 24, 2016

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: All funding pertaining to this study can be found in the Supporting Information file "Supplementaltablesfile.xlsx" under the "Table 11" tab.

Competing Interests: The authors have declared that no competing interests exist.

Copenhagen, Copenhagen, Denmark, **23** Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark, **24** Human Cancer Genetics Program, Spanish National Cancer Research Centre, Madrid, Spain, **25** Centro de Investigación en Red de Enfermedades Raras, Valencia, Spain, **26** Department of Epidemiology, University of California Irvine, Irvine, CA, United States of America, **27** Beckman Research Institute of City of Hope, Duarte, CA, United States of America, **28** Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany, **29** Division of Preventive Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany, **30** German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany, **31** Division of Gynaecology and Obstetrics, Technische Universität München, Munich, Germany, **32** Division of Molecular Gyneco-Oncology, Department of Gynaecology and Obstetrics, University Hospital of Cologne, Cologne, Germany, **33** Center of Familial Breast and Ovarian Cancer, University Hospital of Cologne, Cologne, Germany, **34** Center for Integrated Oncology, University Hospital of Cologne, Cologne, Germany, **35** Center for Molecular Medicine, University Hospital of Cologne, Cologne, Germany, **36** Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany, **37** University of Tübingen, Tübingen, Germany, **38** Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany, **39** Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland, **40** Department of Preventive Medicine, Kyushu University Faculty of Medical Sciences, Fukuoka, Japan, **41** Department of Breast Oncology, Aichi Cancer Center Hospital, Aichi, Japan, **42** Gynaecology Research Unit, Hannover Medical School, Hannover, Germany, **43** Department of Radiation Oncology, Hannover Medical School, Hannover, Germany, **44** Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden, **45** Department of Oncology - Pathology, Karolinska Institutet, Stockholm, Sweden, **46** Cancer Center, Kuopio University Hospital, Kuopio, Finland, **47** Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland, **48** Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio, Finland, **49** Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia, **50** Peter MacCallum Cancer Center, The University of Melbourne, Melbourne, Australia, **51** Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States of America, **52** University Hospital Gashuisberg, Leuven, Belgium, **53** Vesalius Research Center, Leuven, Belgium, **54** Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Leuven, Belgium, **55** Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, **56** University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany, **57** Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS (Istituto Di Ricovero e Cura a Carattere Scientifico) Istituto Nazionale dei Tumori (INT), Milan, Italy, **58** Division of Cancer Prevention and Genetics, Istituto Europeo di Oncologia, Milan, Italy, **59** Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, United States of America, **60** Department of Health Sciences Research, Mayo Clinic, Rochester, MN, United States of America, **61** Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Australia, **62** University of Hawaii Cancer Center, Honolulu, HI, United States of America, **63** Department of Medicine, McGill University, Montreal, Canada, **64** Division of Clinical Epidemiology, Royal Victoria Hospital, McGill University, Montreal, Canada, **65** Cancer Research Initiatives Foundation, Subang Jaya, Selangor, Malaysia, **66** Breast Cancer Research Unit, Cancer Research Institute, University Malaya Medical Centre, KualaLumpur, Malaysia, **67** Department of Genetics, Institute for Cancer Research, Radiumhospitalet, Oslo University Hospital, Oslo University Hospital, Oslo, Norway, **68** Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway, **69** Department of Clinical Molecular Biology, Oslo University Hospital, University of Oslo, Oslo, Norway, **70** Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, United States of America, **71** Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, University of Oulu, Oulu, Finland, **72** Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre NordLab, Oulu, Finland, **73** Department of Oncology, Oulu University Hospital, University of Oulu, Oulu, Finland, **74** Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Canada, **75** Department of Molecular Genetics, University of Toronto, Toronto, Canada, **76** Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, Canada, **77** Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada, **78** Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands, **79** Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, **80** Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, The Netherlands, **81** Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK, **82** Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, **83** Shanghai Center for Disease Control and Prevention, Shanghai, China, **84** Sheffield Cancer Research, Department of Oncology, University of Sheffield, Sheffield, UK, **85** Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK, **86** International Epidemiology Institute, Rockville, MD, United States of America, **87** Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK, **88** Department of Preventive Medicine, Seoul National

University College of Medicine, Seoul, Korea, **89** Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea, **90** Cancer Research Institute, Seoul National University, Seoul, Korea, **91** Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore, **92** Department of Surgery, National University Health System, Singapore, Singapore, **93** Institute of Human Genetics, Pontificia Universidad Javeriana, Bogota, Colombia, **94** Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland, **95** National Cancer Institute, Bangkok, Thailand, **96** International Agency for Research on Cancer, Lyon, France, **97** Molecular Diagnostics Laboratory, IRRP, National Centre for Scientific Research "Demokritos", Athens, Greece, **98** School of Public Health, China Medical University, Taichung, Taiwan, **99** Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, **100** Cancer Center and Department of Surgery, Chung-Ho Memorial Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan, **101** Centre Hospitalier Universitaire de Québec Research Center, Laval University, Québec City, Canada, **102** Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK, **103** Edinburgh Cancer Research UK Centre, University of Edinburgh, Edinburgh, UK

☯ These authors contributed equally to this work.
¶ Membership of the KConFab/AOCS Investigators is listed in the Acknowledgments.
* jonine.figueroa@ed.ac.uk

Abstract

The Cancer Genetic Markers of Susceptibility genome-wide association study (GWAS) originally identified a single nucleotide polymorphism (SNP) rs11249433 at 1p11.2 associated with breast cancer risk. To fine-map this locus, we genotyped 92 SNPs in a 900kb region (120,505,799–121,481,132) flanking rs11249433 in 45,276 breast cancer cases and 48,998 controls of European, Asian and African ancestry from 50 studies in the Breast Cancer Association Consortium. Genotyping was done using iCOGS, a custom-built array. Due to the complicated nature of the region on chr1p11.2: 120,300,000–120,505,798, that lies near the centromere and contains seven duplicated genomic segments, we restricted analyses to 429 SNPs excluding the duplicated regions (42 genotyped and 387 imputed). Per-allelic associations with breast cancer risk were estimated using logistic regression models adjusting for study and ancestry-specific principal components. The strongest association observed was with the original identified index SNP rs11249433 (minor allele frequency (MAF) 0.402; per-allele odds ratio (OR) = 1.10, 95% confidence interval (CI) 1.08–1.13, $P = 1.49 \times 10^{-21}$). The association for rs11249433 was limited to ER-positive breast cancers (test for heterogeneity $P \leq 8.41 \times 10^{-5}$). Additional analyses by other tumor characteristics showed stronger associations with moderately/well differentiated tumors and tumors of lobular histology. Although no significant eQTL associations were observed, in silico analyses showed that rs11249433 was located in a region that is likely a weak enhancer/promoter. Fine-mapping analysis of the 1p11.2 breast cancer susceptibility locus confirms this region to be limited to risk to cancers that are ER-positive.

Introduction

Genome-wide association studies (GWAS) have identified over 90 common genetic variants associated with breast cancer risk [1–19]. A multi-stage GWAS, the Cancer Genetic Markers of Susceptibility (CGEMS) initiative, identified a single nucleotide polymorphism (SNP), rs11249433, associated with breast cancer risk. This SNP is located in the peri-centromeric

region of chromosome 1p11.2, upstream of the *NOTCH2* and *FCGR1B* genes [12]. Further independent analysis confirmed this region as a breast cancer susceptibility locus associated with estrogen receptor (ER) positive but not ER-negative breast cancers [20–22], and more strongly associated with invasive lobular breast cancers than invasive ducal cancers [23]. Two independent meta-analyses on the basis of 15 case-control studies provided data supporting a significant association between rs11249433 and breast cancer among Caucasian populations but did not identify any significant association in Asian and African populations [24, 25].

Fine-scale mapping of the susceptibility regions identified by GWAS has the potential to further narrow down the relevant area of interest, identifying additional risk SNPs, and predicting potential functional mechanisms. Fine-mapping of the 1p11.2 locus among Chinese women (878 cases and 900 controls) identified a novel SNP rs2580520 as a variant significantly associated with breast cancer risk, which was not identified in European women [26]. However, fine-mapping has not been performed at this locus in a large population of multi-ethnic women. The Collaborative Oncological Gene-environment Study (COGS) designed and executed a collaborative genotyping and fine-mapping effort utilizing a custom built iSelect genotyping array (iCOGS) [8]. In this study we fine-mapped the1p11.2 breast cancer susceptibility locus utilizing the data generated through iCOGS, using both genotyped and imputed SNPs from over 50 case-control studies within the Breast Cancer Association Consortium (BCAC). Further, we determined whether the associated SNPs displayed heterogeneity by tumor subtype defined by ER-expression, as well as tumor grade and histology.

Materials and Methods

Study populations

Fifty breast cancer studies participating in the Breast Cancer Association Consortium (BCAC) were included in this analysis. The majority of the included studies were population-based or hospital-based case-control studies that included participants of European ancestry (41 studies), Asian ancestry (9 studies), and African ancestry (2 studies), totaling 45,276 breast cancer cases and 48,998 controls. Study participants were recruited under protocols approved by the Institutional Review Board at each institution, and all subjects provided written informed consent, as previously described [8]. For a list of all approving Institutional Review Boards by study, refer to Table A in S1 File.

SNP selection, genotyping and imputation

Genotyping and quality control (QC) measures used in COGs have been described elsewhere [8]. In brief, excluded were SNPs with call rates of $< 95\%$, with Hardy-Weinberg equilibrium deviation in controls at $P < 1 \times 10^{-7}$ and those with more than 2% of discrepant genotypes in duplicate samples across all COGS consortia. The 900 Kb genomic region for fine-mapping of the 1p11.2 locus (chr1p11.2: 120,300,000–121,185,600; based on build hg19) included all known SNPs correlated ($r^2 > 0.1$) with the index variant rs11249433. In total, 92 genotyped SNPs from the iCOGS array satisfied the initial QC metrics above.

We used imputation in order to estimate the genotypes at variants in the region not typed on the iCOGS array. Imputation was performed using IMPUTE2 [27], separately for each ethnic group. The IMPUTE2-info score and posterior probabilities at each SNP were used to evaluate imputation performance; scores ranged from 1 (high confidence) to 0 (no confidence). Markers with an IMPUTE2-info score < 0.9 or minor allele frequency (MAF) $\leq 3\%$ were excluded from the association analyses as unreliable. Imputed genotypes where the maximum probability was <0.9 were considered unknown. For reference panels we used the International HapMap Project Phase 3 CEU data [28] and the 1000 Genome Project June 2010 release [29].

Based on these reference panels, genotypes at an additional 4,279 SNPs were reliably imputed across the 1p11.2 region.

After reviewing the 1p11.2 genomic region using the UCSC genome browser, we observed that there were several SNPs which mapped to multiple genomic regions due to duplication of genomic segments on both sides of the centromere on chromosome 1 [30, 31]. Therefore, we restricted our analysis to the region of 1p11.2 that excluded the following duplicated genomic segments: chr1:120531871–120697156; chr1:120747157–120936695; chr1:121086699–121133098; chr1:121160483–121222841; chr1:121280229–121351595; chr1:121361172–121418375 and chr1:121418377–121472478. The final analysis was therefore based on 42 genotyped and 387 imputed SNPs across ~210kb of genomic sequence. We also investigated data for a previously described SNP noted to be associated with risk in Asian populations, rs2580520 [26]. Unfortunately, the rs2580520 SNP was not genotyped in the iCOGS effort, is not curated in the 1000 Genomes database that was used for imputation of the 1p11.2 region (www.1000genomes.org/data [32]) and falls within the duplicated region noted above.

Statistical analysis

The LD structure based on the 1000 Genomes CEU data was visualized using the R package snp.plotter [33]. A line graph was constructed displaying likelihood ratio statistics for recombination hotspot using SequenceLDhot software based on the background recombination rates inferred by PHASE v2.1. Physical locations of SNPs were based on hg19 and gene annotation and the LD plot was based on the NCBI RefSeq genes from the UCSC Genome Browser [34].

Standard logistic regression models were used where the common allele was the referent to assess the association of all genotyped and imputed SNPs with breast cancer risk and all analyses (overall and for breast cancer subtypes) used a 1-degree of freedom test (additive model) to estimate the per-allele odds ratio (OR) for the variant allele and corresponding 95% confidence interval (CI) for each SNP. Association analyses were adjusted for study and eight eigenvectors to capture population structure, obtained from principal component analyses [8]. P-values for trend from the Wald test are reported, imputed SNPs were handled using estimated allele dose. To identify SNPs independently associated with breast cancer risk within the 1p11.2 locus we conducted forward stepwise logistic regression analysis separately for each ethnicity (European, Asian and African) conditioning on rs11249433, the top SNP originally identified in CGEMS SNP and the top SNP at this locus in iCOGS. After identifying a novel independent signal, stepwise logistic regression analyses were repeated conditioning on the newly identified SNP rs146784183. Bonferonni adjusted significance was set at $P < 7x10^{-5}$, corrected for 4,371 SNPs.

To determine if there were differences in the associated effects of the independent signals on different subtypes of breast cancer among women of European ancestry, we conducted stratified analyses according to subtypes defined by: 1) tumor histology (ductal/mixed, lobular, other), 2) tumor grade (well-differentiated, moderately-differentiated, poorly-differentiated), and 3) ER status (ER-positive or ER-negative) subtypes. To determine if SNP associations varied significantly between defined subtypes of breast cancer, we performed polytomous logistic regression models, and P-values for heterogeneity were obtained from case-case analysis for tumor subtypes (ER, tumor grade and tumor histology). Meta-analyses were performed using the random effects model to estimate the I^2 statistic and p-value for heterogeneity by study.

In silico functional analysis and eQTL data

To evaluate any possible functional implications of our top-associated SNPs, we assessed *in silico* functional data and expression quantitative trait loci (eQTL). Utilizing the UCSC Genome Browser and HaploReg v3 we reviewed ENCODE data to determine potentially altered

regulatory motifs. RegulomeDB v1.1 was used to query publicly available eQTL data from multiple cell types associated with the identified SNPs and select SNPs significantly correlated to the tag SNP rs11249433.

Results and Discussion

Fine-scale mapping of the 1p11.2 locus

Following quality control and genomic restrictions, a total of 429 SNPs (42 genotyped and 387 imputed) were examined for their association with breast cancer risk. Fig 1 shows the genotyped and imputed SNPs analyzed in European women, plotted against corresponding chromosomal positions within 1p11.2. Gene annotations within this genomic region, including the *NOTCH2* gene, and the degree of linkage disequilibrium between the SNPs, are also shown in Fig 1.

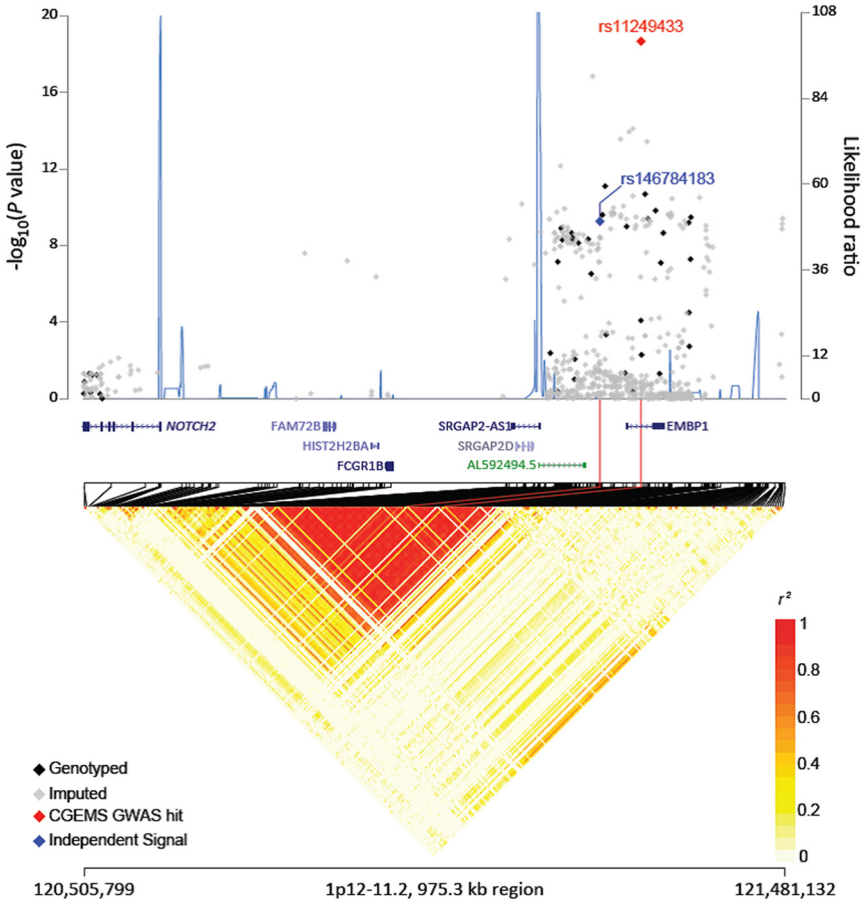

Fig 1. Regional plots of breast cancer association in 1p12-11.2. Regional plot of association result, recombination hotspots and linkage disequilibrium for the 1p12-11.2:120,505,799–121,481,132 breast cancer susceptibility loci. Association result from a trend test in—log10*P*values (y axis, left; red diamond, the top ranked breast cancer associated locus in the region; blue diamond, best conditioned analysis results conditioned on rs11249433; black diamonds, genotyped SNPs; gray diamonds, imputed SNPs) of the SNPs are shown according to their chromosomal positions (x axis). Linkage disequilibrium structure based on the 1000 Genomes CEU data (n = 85) was visualized by snp.plotter software. The line graph shows likelihood ratio statistics (y axis, right) for recombination hotspot by SequenceLDhot software based on the background recombination rates inferred by PHASE v2.1. Physical locations are based on hg19. Gene annotation was based on the NCBI RefSeq genes from the UCSC Genome Browser.

doi:10.1371/journal.pone.0160316.g001

Table 1. Two independent association signals at the 1p11.2 locus: Association results for breast cancer risk among women in BCAC, by ancestry.

Signal	SNP	Position[a]	Cases (N)	Controls (N)	Source[b]	Risk Allele	RAF[c] Cases/Controls	r^{2d}	OR (95% CI)[e]	*P*-trend
European Ancestry										
1	rs11249433	121280613	39,072	42,101	G	G	0.424/0.402	~	1.10 (1.08–1.13)	1.49E-21
									1.09 (1.06–1.11)[f]	6.54E-15
2	rs146784183	121223447	39,072	42,101	I	A	0.906/0.899	0.086	0.88 (0.85–0.91)	2.35E-12
									0.92 (0.88–0.95)[f]	1.27E-05
Asian Ancestry										
1	rs11249433	121280613	5,826	6,643	G	G	0.039/0.036	~	1.19 (1.04–1.36)	0.01
2	rs146784183	121223447	5,826	6,643	I	A	0.860/0.844	0.004	0.89 (0.82–0.96)	0.002

[a]Genomic coordinates are based on hg19.

[b]Source indicates whether the SNP was genotyped (G) or imputed (I) within the data used from the 1000 Genomes Project.

[c]Risk allele frequency (RAF) for cases/controls.

[d]Pair-wise linkage disequilibrium (r2) with the top SNP rs11249433 calculated using iCOGS (n = 84,396) data, for controls only.

[e]Per-allele odds ratios (OR) and 95% confidence intervals (95% CI) were estimated from logistic regression adjusted for study site and 7 principal components in Europeans and 2 principal components in women with Asian ancestry. The common allele was the referent for calculating odds ratio; the G-allele for both rs11249433 and rs146784183.

[f]Odds ratios (OR) and 95% confidence intervals (95% CI) were estimated from logistic regression mutually adjusted for rs146784183 and top SNP (rs11249433) along with study site and 7 principal components.

doi:10.1371/journal.pone.0160316.t001

Breast cancer risk associations at the 1p11.2 locus

Of the 429 SNPs, 136 SNPs were associated with breast cancer risk overall in European women at $P < 5x10^{-8}$ (Table B in S1 File and S1 Fig). The most significant association with breast cancer risk was observed for the previously identified rs11249433 SNP (MAF 0.402; per-G-allele OR = 1.10, 95% CI 1.08–1.13, $P = 1.49 \times 10^{-21}$, Table 1) [12]. To test for the existence of additional independent signals within the 1p11.2 locus, we conducted forward stepwise logistic regression analyses conditioning on the top SNP rs11249433. A second signal was identified corresponding to an imputed SNP rs146784183 (MAF 0.101; per-A-allele OR = 0.88, 95% CI 0.85–0.91, $P = 1.27 \times 10^{-5}$ after adjustment for rs11249433, Table 1). After adjustment for rs11249433, SNP rs146784183 was not strongly correlated with the index SNP ($r^2 = 0.086$), and is located 57 kb telomeric from rs11249433, and closer to the *NOTCH2* gene. Stepwise regression analyses conditioning on both rs146784183 and rs11249433 did not result in the identification of any additional independent signals at this locus (Table C in S1 File). Meta-analyses demonstrated that results were similar across studies for association results seen for both rs11249433 ($I^2 = 0\%$, $P\text{-het}_{study} = 0.844$) and rs146784183 ($I^2 = 6.7\%$, $P\text{-het}_{study} = 0.351$).

Association analysis by estrogen receptor status in European women

We next determined whether risk associations at the 1p11.2 locus varied by estrogen receptor (ER) status; associations observed were limited to ER-positive (rs11249433: per-G-allele OR = 1.12, 95% CI 1.10–1.15, *P*-het $= 9.88 \times 10^{-9}$; rs146784183: per-A-allele OR = 0.86, 95% CI 0.82–0.89, *P-het* $= 8.41 \times 10^{-5}$; Table 2 and Table D in S1 File). Associations for these two SNPs among ER-negative breast cancers were null (rs11249433: per-G-allele OR = 1.00, 95% CI 0.95–1.05, *P* = 0.90; rs146784183: per-A-allele OR = 0.99, 95% CI 0.92–1.06, *P* = 0.68; Table 2 and Table D in S1 File). Meta-analyses stratified by estrogen receptor status demonstrated that results were similar across studies for association results seen for both rs11249433 (ER-positive: $I^2 = 0\%$, $P\text{-het}_{study} = 0.846$) and rs146784183 (ER-positive: $I^2 = 0\%$, $P\text{-het}_{study} = 0.524$).

Table 2. Two independent association signals at the 1p11.2 locus: Association results for breast cancer risk among European women in BCAC, by tumor characteristic.

Tumor Characteristic	Cases (N)	Signal	SNP	OR (95% CI)[a]	*P*-trend	*P*- heterogeneity[b]
ER Status						
ER-positive	6,315	1	rs11249433	1.12 (1.10–1.15)	4.09E-23	
ER-negative	21,610			1.00 (0.95–1.05)	0.90	9.88E-09
ER-positive	6,315	2	rs146784183	0.86 (0.82–0.89)	1.48E-12	
ER-negative	21,610			0.99 (0.92–1.06)	0.68	8.41E-05
Tumor Grade						
Well-differentiated	5,917	1	rs11249433	1.18 (1.14–1.23)	5.63E-17	
Moderately-differentiated	13,561			1.13 (1.10–1.16)	3.88E-17	
Poorly-differentiated	8,784			1.02 (0.98–1.05)	0.27	8.90E-11
Well-differentiated	5,917	2	rs146784183	0.84 (0.77–0.90)	3.23E-06	
Moderately-differentiated	13,561			0.85 (0.81–0.90)	6.45E-09	
Poorly-differentiated	8,784			0.96 (0.90–1.02)	0.20	8.80E-04
Histology						
Ductal/Mixed	22,308	1	rs11249433	1.08 (1.05–1.10)	5.07E-09	
Lobular	3,747			1.28 (1.22–1.35)	1.15E-23	
Other	2,563			1.12 (1.05–1.19)	0.0002	7.60E-11
Ductal/Mixed	22,308	2	rs146784183	0.90 (0.86–0.94)	1.69E-06	
Lobular	3,747			0.81 (0.74–0.89)	8.07E-06	
Other	2,563			0.90 (0.81–1.00)	0.05	0.11

[a]Odds ratios (OR) and 95% confidence intervals (95% CI) were estimated from logistic regression adjusted for study site and 7 principal components. The common allele was the referent for calculating odds ratio; the G-allele for both rs11249433 and rs146784183.

[b]P-heterogeneity tests whether SNP associations differ significantly by tumor characteristic.

doi:10.1371/journal.pone.0160316.t002

Association analysis by tumor grade and histology in European women

Assessment of risk associations by tumor grade showed that SNP rs11249433 was significantly associated with risk for well-differentiated tumors (per-G-allele OR = 1.18, 95% CI 1.14–1.23) and moderately-differentiated tumors (per-G-allele OR = 1.13, 95% CI 1.10–1.16), but not poorly-differentiated tumors (per-G-allele OR = 1.02, 95% CI 0.98–1.05; *P* -het = 8.90×10^{-11},Table 2 and Table E in S1 File). Similarly, SNP rs146784183 showed significant associations for well and moderately-differentiated tumors, but not poorly-differentiated tumors in (*P* -heterogeneity = 8.80×10^{-4},Table 2 and Table E in S1 File). Results were similar when assessing heterogeneity by tumor grade only among ER-positive breast cancers, there were no significant associations for poorly-differentiated ER-positive tumors (Table F in S1 File).

Differential risk associations for rs11249433 was also seen by tumor histology, where associations were strongest for lobular tumors (per-G-allele OR = 1.28, 95% CI 1.22–1.35; *P* -het = 7.60×10^{-11}), and less so for ductal/mixed or other tumor histology (Table 2). Significant risk differences by tumor histology were not observed for SNP rs146784183 (*P* -heterogeneity = 0.11), though the risk reduction associated with this SNP was strongest for lobular tumors (Table 2). Of the 160 genotyped and imputed SNPs found to be significantly associated with lobular breast tumors at a Bonferroni adjusted $P < 7 \times 10^{-5}$, 127 (79%) were also associated with ductal/mixed tumors, and only 30 (19%) of those also associated with tumors of other histology (Table G in S1 File).

Analysis of index SNPs in different ethnic groups

We also examined breast cancer risk associations among participants in the nine case-control studies that included women of Asian ancestry (Table 2, S1 Fig and Table H in S1 File). The degree of linkage disequilibrium between the SNPs in this region was examined using HapMap data (S2 Fig).

The top SNP among European women, rs11249433, was also associated with breast cancer risk among Asian women (per-G-allele OR = 1.19, 95% CI 1.04–1.36; $P = 0.01$, Table 1 and S1 Fig). Although this SNP is rare in this population (MAF = 0.037), the OR was consistent with that in Europeans. SNP rs146784183 was also associated with breast cancer risk among Asian women (per-A-allele OR = 0.89, 95% CI 0.82–0.96; $P = 0.002$, Table 1 and S1 Fig).

The most strongly associated SNP within the Asian population, genotyped SNP rs115775083, was found to be significantly associated with breast cancer risk overall within the Asian population (per-T-allele OR = 1.78, 95% CI 1.43–2.20, $P = 1.52 \times 10^{-7}$, Table H in S1 File). The rs115775083 genotyped SNP is a rare variant among Asian women with a MAF = 0.011. This genotype is also rare among European women (MAF = 0.016) but not associated with breast cancer risk (per-T-allele OR = 0.95, 95% CI 0.88–1.02, $P = 0.15$). SNP rs115775083 is not correlated with the rs11249433 and rs146784183 SNPs identified to be associated with breast cancer risk in European women ($r^2 < 0.01$). Conditioning on the top SNP identified in the women of Asian did not identify any novel signals within the 1p11.2 locus, but did reaffirm SNP rs115775083 as the most significant signal among Asian women (Table I in S1 File). Similar analyses were performed among women with African Ancestry using data from two BCAC studies (N = 378 cases and N = 254 controls). There were no SNPs within the 1p11.2 locus found to be significantly associated with breast cancer risk after adjusting for multiple comparisons (Table J in S1 File and S1 Fig).

In silico functional and eQTL data

SNP rs11249433, was strongly correlated with one other SNP, rs12134101 ($r^2 = 0.943$), which showed a similar association with risk (both for overall and ER-positive breast cancer). All other SNPs were less strongly associated with risk (likelihood ratio < 1:1000 relative to rs11249433), suggesting that one or both SNPs rs11249433 and rs12134101 are likely to be causally implicated in breast cancer risk.

In silico analyses showed that SNP rs11249433 was found to be located within a weak enhancer and weak promoter in myoblasts and leukemia cells, respectively. Also, this SNP was located within a region of DNase I hypersensitivity and histone H3K27 acetylation in multiple cell types including T47D and MCF7 breast cancer cell lines. There were no proposed regulatory motifs altered by SNP rs146784183, and neither rs11249433 nor rs146784183 were found to have any significant eQTL associations.

In this large-scale fine-mapping analysis of nearly 50,000 breast cancer cases and 50,000 controls within the Breast Cancer Association Consortium (BCAC), we found index SNP rs11249433 to be the strongest signal within the 1p11.2 locus associated with breast cancer risk in European women. An additional association signal was identified, rs146784183, that was independent of the index SNP for overall breast cancer risk. Neither signal was found to be significantly associated with breast cancer risk among women with Asian or African ancestry, after adjusting for multiple comparisons. Notably, rs11249433 and rs146784183 displayed significant heterogeneity in risk associations by important tumor characteristics including ER status, tumor grade and histology. Our findings highlight the value of fine-mapping analyses to identify novel risk associations, and the utility of performing large-scale genotyping projects

within varied ethnic populations to aid in narrowing down the genomic area relevant to future functional analyses.

Fine-mapping the 1p11.2 locus was complex due to the proximity to the centromere and the presence of duplicate genomic segments. As such, we employed strict quality control measures to increase our likelihood for finding true association signals. In this study we have identified SNP rs146784183 as a novel independent signal within the 1p11.2 locus among European women. SNP rs146784183 and the index SNP were not correlated ($r^2 = 0.086$), and this newly identified SNP is located about 57 kb telomeric from the index SNP, and closer to the *NOTCH2* gene.

Our findings concur with previous research identifying rs11249433 as a SNP displaying heterogeneity by important tumor characteristics including ER status and histology. Specifically, rs11249433 was found to be more strongly associated with tumors of lobular histology and those that were ER-positive [20–22]. Further, we have recently shown that this SNP was more strongly associated with tumors having low E-cadherin breast tissue expression compared to E-cadherin high tumors [22]. Our current and previous findings for SNP rs11249433 are consistent given that expression of the E-cadherin tumor suppresor protein is frequently lost in tumors of lobular histology.

We did not identify any eQTL signals for either rs11249433 or rs146784183. *In silico* analyses showed that rs11249433 is situated in a DNase I hypersensitive region which contains open chromatin with histone marks, suggestive that this region might be a weak enhancer in some cell types [35]. SNP rs11249433 is located upstream of the *NOTCH2* gene on chromosome 1. The NOTCH signaling pathway has been frequently implicated in breast cancer development though the exact function of *NOTCH2* in this process is not well characterized [36–40]. Interestingly, the *NOTCH2* gene was shown to be associated with super-enhancers, or large clusters of transcriptional enhancers that drive expression of genes that function in the acquisition of hallmark capabilities in cancer [41]. Dysregulation of the NOTCH signaling pathway has been implicated in breast cancer initiation and progression; this pathway is also considered as the target for novel therapeutics [36–40]. Consequently, though rs11249433 is located within a weak enhancer, it is plausible that it participates in transcriptional regulation through the function of a larger super-enhancer that contributes to tumor pathology.

In the current study we did not perform functional analyses, however, in a study of 180 breast tumors Fu and colleagues found that carriers of the risk genotypes of rs11249433 (AG/GG) were associated with increased mRNA expression of the *NOTCH2* gene [20–22]. Further, expression of *NOTCH2* was highest in ER-positive/TP53 wild-type tumors. This study supports the potential regulation of *NOTCH2* gene expression by SNP rs11249433 and in turn, is in line with our observation that this SNP is specific to ER-positive breast tumors. In a separate study of NOTCH2 protein expression in breast cancer, NOTCH2 levels were found to be high in well-differentiated tumors and low in poorly-differentiated tumors [42]. If, as suggested by Fu et al. [21], rs11249433 contributes to the increased expression of *NOTCH2*, the observation by Parr and colleagues that NOTCH2 is highest among well-differentiated tumors, supports our findings for low grade, well-differentiated tumors. However, without direct experimental evidence, it is difficult to determine the functional implications of these SNPs with certainty. While it is possible that these two variants (rs11249433 and rs146784183) are influencing different genes, however, the patterns of association with breast cancer sub-types suggest that they may affect similar biological and/or signaling processes.

Our analyses in a diverse population of women showed that the top association signals found in European women showed similar associations in women of Asian ancestry, although associations were weaker. However, no significant signals were observed among women with African ancestry. These findings support what has been previously shown in multi-ethnic

studies of the 1p11.2 locus [24, 25, 43]. Among Asian women, a rare variant, SNP rs115775083 was found to be the strongest association signal for breast cancer overall. This region of chromosome 1 and its association with breast cancer has been examined among Chinese women. Jiang and colleagues assessed the association of seven tagging SNPs, including rs11249433, within a 277 kb region of 1p11.2 [26]. In the Jiang study, the authors observed borderline significant associations of rs11249433 with breast cancer risk in their population of Chinese women. However, given that this SNP is rare among women with Asian ancestry, the absence of a significant association is likely due to decreased power caused by insufficient numbers of cases harboring the risk allele. rs115775083, the top SNP among Asian women in our population, was not included among the seven SNPs assessed in the Jiang study [26]. We were unable to duplicate the findings from Jiang et al. [26] which identified rs2580520 as a significant association signal among Chinese women. The rs2580520 SNP was not genotyped as part of the iCOGs effort, is not found in the 1000 Genomes Project Phase 1 data [32] which was used for imputation, and maps to a suspected duplicated region. These data illustrate the challenges of genotyping this complex region. Though no significant signals were found among women with African Ancestry, examining the regional plots among European, Asian and African women, association analysis suggests that the relevant area of interest for future studies lies within the interval spanning chr1p11.2: 121,105,799–121,405,799.

The strengths of our study are in analysis of a very large data set, which includes subjects of European, Asian and African ancestry; and availability of detailed genetic and tumor pathology data, which allowed us to refine these risk associations by pathologic subtypes of breast cancer. Moreover, the findings observed in this pooled analysis did not differ significantly by study. Our study was limited by the available genomic information of the 1p11.2 region. However, the genomic map of the peri-centromeric region that harbors our region of interest was significantly improved in the latest build of the reference human genome. Due to this improvement, some genomic gaps were filled and some new pseudogene transcripts were mapped in the region; this could potentially increase SNP coverage and improve fine-mapping quality.

Conclusions

In summary, we showed the 1p11.2 locus is specific for ER-positive breast cancers and provided data to narrow the relevant area of interests for future functional studies, which should provide further insights into the underlying causal SNPs responsible for its association with breast cancer.

Supporting Information

S1 Fig. Regional plots of 1p12-11.2 breast cancer associations in women of European, Asian and African Ancestry. Regional plot of association results for the 1p12-11.2:120,505,799–121,481,132 breast cancer susceptibility loci from women of European (top panel), Asian (middle panel) and African (lower panel) ancestry. Association result from a trend test in—log10*P*values (y axis, left; red diamond, the top ranked breast cancer associated locus among European women; blue diamond, best conditioned analysis results conditioned on rs11249433 among European women) of the SNPs are shown according to their chromosomal positions (x axis). Physical locations are based on hg19.
(DOC)

S2 Fig. LD plots for CEU, Asians and Yoruba (YRI) based on HapMAP version 3. Linkage disequilibrium (LD) plots for (A) women with ancestry from northern and western Europe (CEU), (B) Asian ancestry and (C) Yoruba (YRI) women with West African ancestry based

on HapMAP version 3, chromosome 1: 120505–121481 kb. Index SNP rs11249433 is circled in red.
(DOCX)

S1 File. Table A: List of studies and ethics approvals. Table B: Genotyped and imputed SNPs at the 1p11.2 locus associated with overall breast cancer risk at genome-wide significance ($p < 5x10-8$) among European women in BCAC. Table C: Genotyped and imputed SNPs at the 1p11.2 locus associated breast cancer risk at genome-wide significance ($p < 5x10-8$) after conditioning on SNP rs146784183 among European women in BCAC. Table D: Genotyped and imputed SNPs at the 1p11.2 locus associated with ER-positive breast cancer risk at Bonferroni adjusted significance ($p < 7x10-5$) and the corresponding association with ER-negative breast cancer risk among European women in BCAC. Table E: Genotyped and imputed SNPs at the 1p11.2 locus associated with well differentiated breast cancer risk at Bonferroni adjusted significance ($p < 7x10-5$) and the corresponding association with moderately differentiated and poorly differentiated breast cancer risk among European women in BCAC. Table F: Two independent association signals at the 1p11.2 locus: Association results for breast cancer risk among European women in BCAC, by tumor characteristic. Table G: Genotyped and imputed SNPs at the 1p11.2 locus associated with lobular breast cancer risk at Bonferroni adjusted significance ($p < 7x10-5$) and the corresponding association with ductal or mixed and other breast cancer risk among European women in BCAC. Table H: Top 5 SNPs at the 1p11.2 locus and their association with breast cancer risk among women with Asian ancestry in BCAC. Table I: Results of association analyses after conditioning on the top SNP (rs115775083) identified among women with Asian Ancestry. Table J: Two independent association signals at the 1p11.2 locus, lack of association with breast cancer risk among women with African Ancestry. Table K: Acknowledgements and funding.
(XLSX)

Acknowledgments

Acknowledgments for each participating study are included in Table K in S1 File.

Author Contributions

Conceived and designed the experiments: HNH CCC H. Zhang KY LP K. Michailidou QW JD JLH MCS MKS AB K. Muir A. Lophatananon PAF MWB OF NJ EJS IT BB FM PG TT SEB HF JB AG HA SLN H. Brenner VA A. Meindl RKS H. Brauch UH HN SK K. Matsuo HI TD NVB A. Lindblom SM A. Mannermaa VK GC kConFab/AOCS AHW DVDB A. Smeets H. Zhao JC AR PR MB FJC CV GGG RLM CAH LLM MSG SHT NAMT VK AB WZ M. Shrubsole RW A. Jukkola-Vuorinen ILA JAK PD C. Seynaeve MG KC HD AH JWMM J. Li WL XS AC SSC WB QC M. Shah CL CB P. Harrington DK JC MH KSC MK DT A. Jakubowska J. Lubinski S. Sangrajrang PB S. Slager DY C. Shen MH A. Swerdlow NO JS P. Hall PDPP DFE SJC AMD JDF.

Analyzed the data: HNH CCC H. Zhang KY LP K. Michailidou QW JD DFE SJC AMD JDF.

Contributed reagents/materials/analysis tools: HNH MKB JLH MCS MKS AB K. Muir A. Lophatananon PAF MWB OF NJ EJS IT BB FM PG TT SEB HF JB AG HA SLN H. Brenner VA A. Meindl RKS H. Brauch UH HN SK K. Matsuo HI TD NVB A. Lindblom SM A. Mannermaa VK GC kConFab/AOCS AHW DVDB A. Smeets H. Zhao JC AR PR MB FJC CV GGG RLM CAH LLM MSG SHT NAMT VK AB WZ M. Shrubsole RW A. Jukkola-Vuorinen ILA JAK PD C. Seynaeve MG KC HD AH JWMM JLi WL XS AC SSC WB QC M. Shah

CL CB P. Harrington DK JC MH KSC MK DT A. Jakubowska J. Lubinski S. Sangrajrang PB S. Slager DY C. Shen MH A. Swerdlow NO JS P. Hall PDPP DFE SJC AMD JDF.

Wrote the paper: HNH CCC H. Zhang KY LP K. Michailidou MKB QW JD JLH MCS MKS AB K. Muir A. Lophatananon PAF MWB OF NJ EJS IT BB FM PG TT SEB HF JB AG HA SLN H. Brenner VA A. Meindl RKS H. Brauch UH HN SK K. Matsuo HI TD NVB A. Lindblom SM A. Mannermaa VK GC kConFab/AOCS AHW DVDB A. Smeets H. Zhao JC AR PR MB FJC CV GGG RLM CAH LLM MSG SHT NAMT VK AB WZ M. Shrubsole RW A. Jukkola-Vuorinen ILA JAK PD C. Seynaeve MG KC HD AH JWMM JLi WL XS AC SSC WB QC M. Shah CL CB P. Harrington DK JC MH KSC MK DT A. Jakubowska J. Lubinski S. Sangrajrang PB S. Slager DY C. Shen MH A. Swerdlow NO JS P. Hall PDPP DFE SJC AMD JDF.

References

1. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nature genetics. 2009; 41(5):585–90. Epub 2009/03/31. doi: 10.1038/ng.354 PMID: 19330027; PubMed Central PMCID: PMC2748125.
2. Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. Nature genetics. 2010; 42(10):885–92. Epub 2010/09/21. doi: 10.1038/ng.669 PMID: 20852631; PubMed Central PMCID: PMC3130795.
3. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007; 447(7148):1087–93. Epub 2007/05/29. doi: 10.1038/nature05887 PMID: 17529967; PubMed Central PMCID: PMC2714974.
4. Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. Journal of the National Cancer Institute. 2011; 103(5):425–35. doi: 10.1093/jnci/djq563 PMID: 21263130.
5. Ghoussaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, Dicks E, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. Nature genetics. 2012; 44(3):312–8. doi: 10.1038/ng.1049 PMID: 22267197; PubMed Central PMCID: PMC3653403.
6. Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nature genetics. 2011; 43(12):1210–4. doi: 10.1038/ng.985 PMID: 22037553; PubMed Central PMCID: PMC3279120.
7. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nature genetics. 2007; 39(7):870–4. doi: 10.1038/ng2075 PMID: 17529973; PubMed Central PMCID: PMC3493132.
8. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nature genetics. 2013; 45(4):353–61, 61e1–2. Epub 2013/03/29. doi: 10.1038/ng.2563 PMID: 23535729.
9. Siddiq A, Couch FJ, Chen GK, Lindstrom S, Eccles D, Millikan RC, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. Human molecular genetics. 2012. Epub 2012/09/15. doi: 10.1093/hmg/dds381 PMID: 22976474.
10. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nature genetics. 2007; 39(7):865–9. Epub 2007/05/29. doi: 10.1038/ng2064 PMID: 17529974.
11. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nature genetics. 2008; 40(6):703–6. Epub 2008/04/29. doi: 10.1038/ng.131 PMID: 18438407.
12. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nature genetics. 2009; 41(5):579–84. Epub 2009/03/31. ng.353 [pii] doi: 10.1038/ng.353 PMID: 19330030; PubMed Central PMCID: PMC2928646.
13. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nature genetics. 2009; 41(3):324–8. Epub 2009/02/17. doi: 10.1038/ng.318 PMID: 19219042; PubMed Central PMCID: PMC2754845.

14. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nature genetics. 2010; 42(6):504–7. Epub 2010/05/11. doi: 10.1038/ng.586 PMID: 20453838; PubMed Central PMCID: PMC3632836.

15. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nature genetics. 2013; 45 (4):392–8, 8e1–2. Epub 2013/03/29. ng.2561 [pii] doi: 10.1038/ng.2561 PMID: 23535733; PubMed Central PMCID: PMC3771695.

16. Bojesen SE, Pooley KA, Johnatty SE, Beesley J, Michailidou K, Tyrer JP, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nature genetics. 2013; 45(4):371–84, 84e1–2. Epub 2013/03/29. ng.2566 [pii] doi: 10.1038/ng.2566 PMID: 23535731; PubMed Central PMCID: PMC3670748.

17. Purrington KS, Slager S, Eccles D, Yannoukakos D, Fasching PA, Miron P, et al. Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. Carcinogenesis. 2014; 35(5):1012–9. Epub 2013/12/12. bgt404 [pii] doi: 10.1093/carcin/bgt404 PMID: 24325915; PubMed Central PMCID: PMC4004200.

18. Cai Q, Zhang B, Sung H, Low SK, Kweon SS, Lu W, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. Nature genetics. 2014; 46(8):886–90.

19. Long J, Cai Q, Sung H, Shi J, Zhang B, Choi JY, et al. Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. PLoS Genet. 2012; 8(2):e1002532.

20. Figueroa JD, Garcia-Closas M, Humphreys M, Platte R, Hopper JL, Southey MC, 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. Human molecular genetics. 2011; 20(23):4693–706. Epub 2011/08/20. doi: 10.1093/hmg/ddr368 PMID: 21852249; PubMed Central PMCID: PMC3209823.

21. Fu YP, Edvardsen H, Kaushiva A, Arhancet JP, Howe TM, Kohaar I, et al. NOTCH2 in breast cancer: association of SNP rs11249433 with gene expression in ER-positive breast tumors without TP53 mutations. Molecular cancer. 2010; 9:113. Epub 2010/05/21. doi: 10.1186/1476-4598-9-113 PMID: 20482849; PubMed Central PMCID: PMC2887795.

22. Horne HN, Sherman ME, Garcia-Closas M, Pharoah PD, Blows FM, Yang XR, et al. Breast cancer susceptibility risk associations and heterogeneity by E-cadherin tumor tissue expression. Breast cancer research and treatment. 2014; 143(1):181–7. doi: 10.1007/s10549-013-2771-z PMID: 24292867.

23. Sawyer E, Roylance R, Petridis C, Brook MN, Nowinski S, Papouli E, et al. Genetic predisposition to in situ and invasive lobular carcinoma of the breast. PLoS Genet. 2014; 10(4):e1004285. doi: 10.1371/journal.pgen.1004285 PMID: 24743323; PubMed Central PMCID: PMC3990493.

24. Chen Q, Shi R, Liu W, Jiang D. Assessing interactions between the association of common genetic variant at 1p11 (rs11249433) and hormone receptor status with breast cancer risk. PloS one. 2013; 8(8): e72487. doi: 10.1371/journal.pone.0072487 PMID: 23977306; PubMed Central PMCID: PMC3745461.

25. Wu S, Cai J, Wang H, Zhang H, Yang W. Association between 1p11-rs11249433 polymorphism and breast cancer susceptibility: evidence from 15 case-control studies. PloS one. 2013; 8(8):e72526. doi: 10.1371/journal.pone.0072526 PMID: 23977314; PubMed Central PMCID: PMC3744559.

26. Jiang Y, Shen H, Liu X, Dai J, Jin G, Qin Z, et al. Genetic variants at 1p11.2 and breast cancer risk: a two-stage study in Chinese women. PloS one. 2011; 6(6):e21563. doi: 10.1371/journal.pone.0021563 PMID: 21738711; PubMed Central PMCID: PMC3124527.

27. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5(6):e1000529. doi: 10.1371/journal.pgen.1000529 PMID: 19543373; PubMed Central PMCID: PMC2689936.

28. International HapMap C, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010; 467(7311):52–8. doi: 10.1038/nature09298 PMID: 20811451; PubMed Central PMCID: PMC3173859.

29. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467(7319):1061–73. doi: 10.1038/nature09534 PMID: 20981092; PubMed Central PMCID: PMC3042601.

30. Musumeci L, Arthur JW, Cheung FS, Hoque A, Lippman S, Reichardt JK. Single nucleotide differences (SNDs) in the dbSNP database may lead to errors in genotyping and haplotyping studies. Hum Mutat. 2010; 31(1):67–73. Epub 2009/10/31. doi: 10.1002/humu.21137 PMID: 19877174; PubMed Central PMCID: PMC2797835.

31. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, et al. Diversity of human copy number variation and multicopy genes. Science. 2010; 330(6004):641–6. Epub 2010/10/30. 330/6004/641 [pii] doi: 10.1126/science.1197005 PMID: 21030649; PubMed Central PMCID: PMC3020103.

32. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015; 526(7571):68–74. doi: 10.1038/nature15393 PMID: 26432245; PubMed Central PMCID: PMC4750478.

33. Luna A, Nicodemus KK. snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. Bioinformatics. 2007; 23(6):774–6. doi: 10.1093/bioinformatics/btl657 PMID: 17234637.

34. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome research. 2002; 12(6):996–1006. doi: 10.1101/gr.229102. Article published online before print in May 2002. PMID: 12045153; PubMed Central PMCID: PMC186604.

35. Corradin O, Scacheri PC. Enhancer variants: evaluating functions in common disease. Genome Med. 2014; 6(10):85.

36. Bouras T, Pal B, Vaillant F, Harburg G, Asselin-Labat ML, Oakes SR, et al. Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. Cell stem cell. 2008; 3(4):429–41. Epub 2008/10/23. doi: 10.1016/j.stem.2008.08.001 PMID: 18940734.

37. Hirose H, Ishii H, Mimori K, Ohta D, Ohkuma M, Tsujii H, et al. Notch pathway as candidate therapeutic target in Her2/Neu/ErbB2 receptor-negative breast tumors. Oncology reports. 2010; 23(1):35–43. PMID: 19956862.

38. Li L, Zhao F, Lu J, Li T, Yang H, Wu C, et al. Notch-1 signaling promotes the malignant features of human breast cancer through NF-kappaB activation. PloS one. 2014; 9(4):e95912. doi: 10.1371/journal.pone.0095912 PMID: 24760075; PubMed Central PMCID: PMC3997497.

39. Suman S, Das TP, Damodaran C. Silencing NOTCH signaling causes growth arrest in both breast cancer stem cells and breast cancer cells. British journal of cancer. 2013; 109(10):2587–96. doi: 10.1038/bjc.2013.642 PMID: 24129237; PubMed Central PMCID: PMC3833225.

40. Wu F, Stutzman A, Mo YY. Notch signaling and its role in breast cancer. Frontiers in bioscience: a journal and virtual library. 2007; 12:4370–83. PMID: 17485381.

41. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155(4):934–47. Epub 2013/10/15. S0092-8674(13)01227-0 [pii] doi: 10.1016/j.cell.2013.09.053 PMID: 24119843; PubMed Central PMCID: PMC3841062.

42. Parr C, Watkins G, Jiang WG. The possible correlation of Notch-1 and Notch-2 with clinical outcome and tumour clinicopathological parameters in human breast cancer. Int J Mol Med. 2004; 14(5):779–86. Epub 2004/10/20. PMID: 15492845.

43. Long J, Zhang B, Signorello LB, Cai Q, Deming-Halverson S, Shrubsole MJ, et al. Evaluating genome-wide association study-identified breast cancer risk variants in African-American women. PloS one. 2013; 8(4):e58350. doi: 10.1371/journal.pone.0058350 PMID: 23593120; PubMed Central PMCID: PMC3620157.