Please cite this article in press as: Darabi et al., Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate *NRBF2* Expression, The American Journal of Human Genetics (2015), http://dx.doi.org/10.1016/j.ajhg.2015.05.002

ARTICLE

Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate *NRBF2* Expression

Hatef Darabi,[1,119] Karen McCue,[2,119] Jonathan Beesley,[2] Kyriaki Michailidou,[3] Silje Nord,[4,5] Siddhartha Kar,[6] Keith Humphreys,[1] Deborah Thompson,[3] Maya Ghoussaini,[6] Manjeet K. Bolla,[3] Joe Dennis,[3] Qin Wang,[3] Sander Canisius,[7] Christopher G. Scott,[8] Carmel Apicella,[9] John L. Hopper,[9] Melissa C. Southey,[10] Jennifer Stone,[11] Annegien Broeks,[7] Marjanka K. Schmidt,[7] Rodney J. Scott,[12,13] Artitaya Lophatananon,[14] Kenneth Muir,[14,15] Matthias W. Beckmann,[16] Arif B. Ekici,[17] Peter A. Fasching,[16,18] Katharina Heusinger,[16] Isabel dos-Santos-Silva,[19] Julian Peto,[19] Ian Tomlinson,[20] Elinor J. Sawyer,[21] Barbara Burwinkel,[22,23] Frederik Marme,[22,24] Pascal Guénel,[25,26] Thérèse Truong,[25,26] Stig E. Bojesen,[27,28,29] Henrik Flyger,[30] Javier Benitez,[31,32,33] Anna González-Neira,[31]

(Author list continued on next page)

Genome-wide association studies have identified SNPs near *ZNF365* at 10q21.2 that are associated with both breast cancer risk and mammographic density. To identify the most likely causal SNPs, we fine mapped the association signal by genotyping 428 SNPs across the region in 89,050 European and 12,893 Asian case and control subjects from the Breast Cancer Association Consortium. We identified four independent sets of correlated, highly trait-associated variants (iCHAVs), three of which were located within *ZNF365*. The most strongly risk-associated SNP, rs10995201 in iCHAV1, showed clear evidence of association with both estrogen receptor (ER)-positive (OR = 0.85 [0.82–0.88]) and ER-negative (OR = 0.87 [0.82–0.91]) disease, and was also the SNP most strongly associated with percent mammographic density. iCHAV2 (lead SNP, chr10: 64,258,684:D) and iCHAV3 (lead SNP, rs7922449) were also associated with ER-positive (OR = 0.93 [0.91–0.95] and OR = 1.06 [1.03–1.09]) and ER-negative (OR = 0.95 [0.91–0.98] and OR = 1.08 [1.04–1.13]) disease. There was weaker evidence for iCHAV4, located 5′ of *ADO*, associated only with ER-positive breast cancer (OR = 0.93 [0.90–0.96]). We found 12, 17, 18, and 2 candidate causal SNPs for breast cancer in iCHAVs 1–4, respectively. Chromosome conformation capture analysis showed that iCHAV2 interacts with the *ZNF365* and *NRBF2* (more than 600 kb away) promoters in normal and cancerous breast epithelial cells. Luciferase assays did not identify SNPs that affect transactivation of *ZNF365,* but identified a protective haplotype in iCHAV2, associated with silencing of the *NRBF2* promoter, implicating this gene in the etiology of breast cancer.

Introduction

Breast cancer is one of the most commonly occurring epithelial malignancies in women, with an estimated 1.7 million new cases and more than 520,000 deaths annually worldwide.[1,2] Familial aggregation and twin studies have shown a substantial contribution of inherited susceptibility to breast cancer.[3] Genome-wide association studies (GWASs) provide a powerful approach to identify common disease alleles. In a two-stage genome-wide association study conducted in European descendants, Turnbull et al.[4] identified five new susceptibility loci.

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden; [2]Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia; [3]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK; [4]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, 0310 Oslo, Norway; [5]K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Kirkeveien 166, 0450 Oslo, Norway; [6]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK; [7]Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, 1066 CX Amsterdam, the Netherlands; [8]Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; [9]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC 3010, Australia; [10]Department of Pathology, The University of Melbourne, Melbourne, VIC 3010, Australia; [11]Centre for Genetic Origins of Health and Disease, University of Western Australia, Crawley, WA 6009, Australia; [12]Discipline of Medical Genetics, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, Newcastle, NSW 2308, Australia; [13]Division of Molecular Medicine, Pathology North, John Hunter Hospital and The Hunter Medical Research Institute, Newcastle, NSW 2305, Australia; [14]Division of Health Sciences, Warwick Medical School, Warwick University, Coventry CV4 7AL, UK; [15]Institute of Population Health, University of Manchester, Manchester M13 9PL, UK; [16]University Breast Center Franconia, Department of Gynecology and Obstetrics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, 91054 Erlangen, Germany; [17]Institute of Human Genetics, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nuremberg, Comprehensive Cancer Center Erlangen-EMN, 91054 Erlangen, Germany; [18]Department of Medicine Division of Hematology and Oncology, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA; [19]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK; [20]Wellcome Trust Centre for Human Genetics and Oxford Biomedical Research Centre, University of Oxford, Oxford OX3 7BN, UK; [21]Research Oncology, Division of Cancer Studies, King's College London, Guy's Hospital, London SE1 9RT, UK; [22]National Center for Tumor Diseases, University of Heidelberg, 69120 Heidelberg, Germany; [23]Division of Molecular Genetic

(Affiliations continued on next page)

Please cite this article in press as: Darabi et al., Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate *NRBF2* Expression, The American Journal of Human Genetics (2015), http://dx.doi.org/10.1016/j.ajhg.2015.05.002

Hoda Anton-Culver,[34] Susan L. Neuhausen,[35] Volker Arndt,[36] Hermann Brenner,[36,37,38] Christoph Engel,[39] Alfons Meindl,[40] Rita K. Schmutzler,[41,42,43] German Consortium of Hereditary Breast and Ovarian Cancer, Norbert Arnold,[44] Hiltrud Brauch,[38,45,46] Ute Hamann,[47] Jenny Chang-Claude,[48] Sofia Khan,[49] Heli Nevanlinna,[49] Hidemi Ito,[50] Keitaro Matsuo,[51] Natalia V. Bogdanova,[52] Thilo Dörk,[53] Annika Lindblom,[54] Sara Margolin,[55] kConFab/AOCS Investigators, Veli-Matti Kosma,[56,57,58] Arto Mannermaa,[56,57,58] Chiu-chen Tseng,[59] Anna H. Wu,[59] Giuseppe Floris,[60] Diether Lambrechts,[61,62] Anja Rudolph,[48] Paolo Peterlongo,[63] Paolo Radice,[64] Fergus J. Couch,[65] Celine Vachon,[8] Graham G. Giles,[9,66] Catriona McLean,[67] Roger L. Milne,[9,66] Pierre-Antoine Dugué,[66] Christopher A. Haiman,[59] Gertraud Maskarinec,[68] Christy Woolcott,[69] Brian E. Henderson,[59] Mark S. Goldberg,[70,71] Jacques Simard,[72] Soo H. Teo,[73,74] Shivaani Mariapun,[73,74] Åslaug Helland,[75] Vilde Haakensen,[76] Wei Zheng,[77] Alicia Beeghly-Fadiel,[77] Rulla Tamimi,[78,79,80] Arja Jukkola-Vuorinen,[81] Robert Winqvist,[82,83] Irene L. Andrulis,[84,85] Julia A. Knight,[86,87] Peter Devilee,[88,89] Robert A.E.M. Tollenaar,[90] Jonine Figueroa,[91] Montserrat García-Closas,[92,93] Kamila Czene,[1] Maartje J. Hooning,[94] Madeleine Tilanus-Linthorst,[95]

(Author list continued on next page)

Epidemiology, German Cancer Research Centre (DZFK), 69047 Heidelberg, Germany; [24]Department of Obstetrics and Gynecology, University of Heidelberg, 69120 Heidelberg, Germany; [25]University Paris-Sud, UMRS 1018, 94807 Villejuif, France; [26]INSERM (National Institute of Health and Medical Research), CESP (Center for Research in Epidemiology and Population Health), U1018, Environmental Epidemiology of Cancer, 94807 Villejuif, France; [27]Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, 2730 Herlev, Denmark; [28]Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, 2730 Herlev, Denmark; [29]Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark; [30]Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, 2730 Herlev, Denmark; [31]Human Genotyping (CEGEN) Unit, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid 28029, Spain; [32]Human Genetics Group, Spanish National Cancer Centre (CNIO), Madrid 28029, Spain; [33]Biomedical Network on Rare Diseases (CIBERER), Madrid 28029, Spain; [34]Department of Epidemiology, University of California Irvine, Irvine, CA 92697, USA; [35]Beckman Research Institute of City of Hope, Duarte, CA 91010, USA; [36]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [37]Division of Preventive Oncology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [38]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [39]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, 04107 Leipzig, Germany; [40]Division of Gynaecology and Obstetrics, Technische Universität München, 81675 Munich, Germany; [41]Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne 50932, Germany; [42]Center for Hereditary Breast and Ovarian Cancer, Medical Faculty, University Hospital Cologne, Cologne 50937, Germany; [43]Center for Integrated Oncology (CIO), Medical Faculty, University Hospital Cologne, Cologne 50937, Germany; [44]Department of Gynaecology and Obstetrics, University Hospital of Schleswig-Holstein, Campus Kiel, Christian-Albrechts University Kiel, 24105 Kiel, Germany; [45]Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, 70376 Stuttgart, Germany; [46]University of Tübingen, 72074 Tübingen, Germany; [47]Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [48]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [49]Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, 00029 HUS, Finland; [50]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Aichi 464-8681, Japan; [51]Department of Preventive Medicine, Kyushu University Faculty of Medical Sciences, Fukuoka 812-8582, Japan; [52]Department of Radiation Oncology, Hannover Medical School, 30625 Hannover, Germany; [53]Gynaecology Research Unit, Hannover Medical School, 30625 Hannover, Germany; [54]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm 17177, Sweden; [55]Department of Oncology - Pathology, Karolinska Institutet, Stockholm 17177, Sweden; [56]School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio 70211, Finland; [57]Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, Kuopio 70211, Finland; [58]Cancer Center of Eastern Finland, University of Eastern Finland, Kuopio 70211, Finland; [59]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA; [60]University Hospital Gashuisberg, 3000 Leuven, Belgium; [61]Vesalius Research Center (VRC), VIB, Leuven 3000, Belgium; [62]Laboratory for Translational Genetics, Department of Oncology, University of Leuven, Leuven 3000, Belgium; [63]IFOM, the FIRC (Italian Foundation for Cancer Research) Institute of Molecular Oncology, 20139 Milan, Italy; [64]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), 20133 Milan, Italy; [65]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA; [66]Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, VIC 3053, Australia; [67]Anatomical Pathology, The Alfred Hospital, Melbourne, VIC 3004, Australia; [68]Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI 96813, USA; [69]Departments of Obstetrics & Gynaecology and Pediatrics, Dalhousie University, Halifax, NS B3H 4R2, Canada; [70]Department of Medicine, McGill University, Montreal, QC H3A 1W7, Canada; [71]Division of Clinical Epidemiology, McGill University Health Centre, Royal Victoria Hospital, Montreal, QC H3H 2R9, Canada; [72]Centre Hospitalier Universitaire de Québec Research Center and Laval University, Quebec City, QC G1V 4G2, Canada; [73]Breast Cancer Research Unit, University Malaya Cancer Research Institute, University Malaya Medical Centre (UMMC), 50603 Kuala Lumpur, Malaysia; [74]Cancer Research Initiatives Foundation, Sime Darby Medical Centre, 47500 Subang Jaya, Malaysia; [75]Department of Oncology, Oslo University Hospital, Radiumhospitalet, 0310 Oslo, Norway; [76]Department of Genetics, Oslo University Hospital, Radiumhospitalet, 0310 Oslo, Norway; [77]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN 37203, USA; [78]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA; [79]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA; [80]Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA; [81]Department of Oncology, Oulu University Hospital and University of Oulu, 90014 Oulu, Finland; [82]Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Chemistry and Biocenter Oulu, University of Oulu, 90014 Oulu, Finland; [83]Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre NordLab, 90220 Oulu, Finland; [84]Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada; [85]Departments of Molecular Genetics and Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON M5S 3G3, Canada; [86]Prosserman Centre for Health Research, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada; [87]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5S 3G3, Canada; [88]Department of Human Genetics, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands; [89]Department of Pathology, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands; [90]Department of Surgical Oncology, Leiden University Medical Center, 2333 ZC Leiden, the Netherlands; [91]Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD 20850, USA; [92]Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton

(Affiliations continued on next page)

Please cite this article in press as: Darabi et al., Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate *NRBF2* Expression, The American Journal of Human Genetics (2015), http://dx.doi.org/10.1016/j.ajhg.2015.05.002

Jingmei Li,[96] Yu-Tang Gao,[97] Xiao-Ou Shu,[77] Angela Cox,[98] Simon S. Cross,[99] Robert Luben,[100] Kay-Tee Khaw,[101] Ji-Yeob Choi,[102,103] Daehee Kang,[102,103,104] Mikael Hartman,[105] Wei Yen Lim,[105,106] Maria Kabisch,[47] Diana Torres,[47,107] Anna Jakubowska,[108] Jan Lubinski,[108] James McKay,[109] Suleeporn Sangrajrang,[110] Amanda E. Toland,[111] Drakoulis Yannoukakos,[112] Chen-Yang Shen,[113,114] Jyh-Cherng Yu,[115] Argyrios Ziogas,[34] Minouk J. Schoemaker,[92] Anthony Swerdlow,[92,116] Anne-Lise Borresen-Dale,[4,117] Vessela Kristensen,[4,117,118] Juliet D. French,[2] Stacey L. Edwards,[2] Alison M. Dunning,[6] Douglas F. Easton,[3,6] Per Hall,[1] and Georgia Chenevix-Trench,[2,*]

One of the identified SNPs, rs10995190 (combined $p = 5.1 \times 10^{-15}$), lies within intron 4 of *ZNF365* (MIM: 607818) at 10q21.2. Analysis by the Breast Cancer Association Consortium (BCAC) confirmed this association ($p = 1.3 \times 10^{-36}$)[5] in those of European ancestry, with a non-significant association among Asians where this variant is rare.[6] Instead, Cai et al.[7] identified a common SNP, rs10822013 located 26.7 kb upsteam, associated with breast cancer risk among East Asians. Further analysis by the Consortium of Investigators of Modifiers of *BRAC1* and *BRACA2* (CIMBA) showed that rs10995190 is associated with breast cancer risk among *BRCA2* carriers,[8] particularly for estrogen-receptor-positive breast cancer. This 10q21.2 locus was the first to be associated with percent mammographic density (PD) (combined $p = 9.6 \times 10^{-10}$)[9] and also shows a possible association with breast size.[10] Varghese et al.[11] showed that rs10995190 and rs10509168 ($r^2 = 0.13$) are both associated with mammographic density, and through analysis of polygenic risk scores found that PD and breast cancer have a shared genetic basis.

These data indicate that 10q21.2 is an important susceptibility region for both breast cancer and mammographic density and must harbor one or more SNPs causally related to these phenotypes. In an attempt to identify the most likely causal SNPs underlying these associations, we assessed 428 SNPs across the 10q21.2 region, applying multiple analyses aimed at exploring the target genes and functional basis of the associations with breast cancer risk and mammographic density.

Material and Methods

Genetic Mapping

Tagging Strategy for the Fine-Scale Mapping

We identified the fine-mapping region by taking the furthest SNPs upstream and downstream with minor allele frequency (MAF) > 2% and detectable correlation ($r^2 > 0.1$) with rs10995190, based on the 1000 Genomes Project European population (March 2010 Pilot version; data from 60 CEU individuals). We also selected SNPs that tagged all remaining SNPs in the 560 kb interval with $r^2 > 0.9$. With this strategy, we selected for inclusion on the iCOGS array (see below) a total of 440 SNPs, between positions 64205327 and 64765654 (NCBI build 37 assembly), that had an Illumina designability score (DS) above 0.9. Of these, 428 were successfully genotyped on the iCOGS array and passed QC filters.

iCOGS Genotyping and Imputation

Case and control samples were drawn from 50 studies participating in the BCAC, of which 41 were from populations of predominantly European ancestry and 9 from populations of Asian ancestry, as described previously.[5] We performed iCOGS genotyping in four centers, as part of the Collaborative Oncological Gene-Environment Study (COGS). All BCAC studies were approved by the relevant local ethics committees, as described previously,[5] and proper informed consent was obtained from all participants. We used the genotype data from the 428 SNPs that passed quality control to impute genotypes at all additional known SNPs in the interval, using IMPUTE v.2.0 (IMPUTE2) and the 1000 Genomes Project data (March 2012 version) as a reference panel.[12] Rather than preselecting a reference population, we followed the approach of Howie et al.[13] and used a multi-population reference panel. IMPUTE2 was applied with default parameters and effective population size (N_e) of 20,000.

SM2 5NG, UK; [93]Division of Cancer Studies, Breakthrough Breast Cancer Research Centre, London SW3 6JB, UK; [94]Department of Medical Oncology, Erasmus MC Cancer Institute, 3008 AE Rotterdam, the Netherlands; [95]Department of Surgical Oncology, Erasmus University Medical Center, 3075 EA Rotterdam, the Netherlands; [96]Human Genetics Division, Genome Institute of Singapore, Singapore 138672, Singapore; [97]Department of Epidemiology, Shanghai Cancer Institute, Xuhui, Shanghai 200031, China; [98]Sheffield Cancer Research Centre, Department of Oncology, University of Sheffield, Sheffield S10 2RX, UK; [99]Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield S10 2HQ, UK; [100]Clinical Gerontology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK; [101]Department of Public Health and Primary Care, University of Cambridge, Cambridge CB2 1TN, UK; [102]Department of Biomedical Science, Seoul National University College of Medicine, Seoul 110-799, Korea; [103]Department of Surgery, Seoul National University College of Medicine, Seoul 110-799, Korea; [104]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul 110-799, Korea; [105]Saw Swee Hock School of Public Health, National University of Singapore and Department of Surgery, National University Health System, Singapore 117597, Singapore; [106]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore 119228, Singapore; [107]Institute of Human Genetics, Pontificia Universidad Javeriana, Bogota 12362, Columbia; [108]Department of Genetics and Pathology, Pomeranian Medical University, 70-115 Szczecin, Poland; [109]International Agency for Research on Cancer, 69372 Lyon, CEDEX 08, France; [110]National Cancer Institute, Bangkok 10400, Thailand; [111]Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA; [112]Molecular Diagnostics Laboratory, IRRP, National Centre for Scientific Research "Demokritos," Aghia Paraskevi Attikis, 153 10 Athens, Greece; [113]Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan; [114]School of Public Health, China Medical University, Taichung 40402, Taiwan; [115]Department of Surgery, Tri-Service General Hospital, Taipei 114, Taiwan; [116]Division of Breast Cancer Research, Institute of Cancer Research, Sutton SM2 5NG, UK; [117]Institute of Clinical Medicine, University of Oslo (UiO), Oslo 0316, Norway; [118]Department of Clinical Molecular Biology (EpiGen), University of Oslo (UiO), Oslo 0316, Norway

[119]These authors contributed equally to this work

*Correspondence: georgia.trench@qimr.edu.au

http://dx.doi.org/10.1016/j.ajhg.2015.05.002.

Mammographic Density

Mammographic density measurements were available for 6,886 women from nine BCAC studies that were also part of the Marker Of DEnsity (MODE) consortium[6] and DENSNP consortium[14] (Table S1). Covariate data were obtained through self-administered postal questionnaires, in-person interviews, or telephone interviews, and anthropometric variables were self-reported or measured by trained staff (Table S2 of Vachon et al.[14]). Seven studies estimated density (absolute and percent dense area) with the CUMULUS program,[15] and two of the studies, the BBCC and NBCS studies, used the Madena software.[16]

Statistical Analysis

For each SNP, we estimated a per-allele log-odds ratio (OR) and standard error using logistic regression, including principal components and per-study fixed-effects to adjust for study-specific differences in allele frequency, as described previously.[5] Analyses were carried out separately for Europeans and Asians. We estimated genetic main effects for ER-positive and ER-negative breast cancer using logistic regression and restricting the cases to a specific subtype. We evaluated heterogeneity of association by tumor subtype in a case-only analysis, treating subtype status as the dependent variable. We derived the p values for association by means of a likelihood-ratio test (one degree of freedom); all tests were two-sided. To identify the most parsimonious model, we identified all SNPs with $p < 10^{-4}$ and MAF $\geq$ 2% in the single SNP analysis and included these in forward selection regression analyses, utilizing the step function in R with penalty term set to 10.[17] To account for uncertainty in the data resulting from the imputation process, we conducted analysis by regressing on the allele dosage for each genotype. We estimated haplotype-specific ORs using an EM algorithm implemented in the haplo.stats package in R. For this analysis we used the most probable genotypes and included study and principal components as covariates. We grouped haplotypes with a frequency < 0.01 together into one subgroup of rare haplotypes.

We assessed individual SNP associations with percentage density (PD), dense area (DA), and non-dense area (nDA) via linear regression. Because the distributions of estimated PD, DA, and nDA were positively skewed (skewness = 1.31, 2.32, and 1.02, respectively), each phenotype was square-root transformed. This transformation reduces skewness and has previously been shown to generate variables that are approximately normally distributed.[18] In addition to study and principal components, we included age, body mass index (BMI), postmenopausal hormone replacement therapy (HRT), mammographic view, menopausal status, and case-control status as covariates in each model. Age and BMI were treated as continuous variables, and use of postmenopausal hormones at time of mammogram (0, never; 1, stopped prior to mammogram date; 2, current use at date of mammogram; 9, unknown), menopausal status at time of mammogram (0, postmenopausal; 1, premenopausal; 2, perimenopausal; 9, unknown), and case-control status were treated as categorical variables in the regression model. We also included mammographic view (1, medio-lateral oblique [MLO] view; 2, craniocaudal [CC] view) and treated it as categorical variable because it has been shown that the percent density measurements from the MLO view are systematically lower than those from the CC view.[9] The mean percentage density across all studies was 17 from the MLO view as compared to 25 from the CC view.

Expression Quantitative Trait Locus Analysis

We examined the associations of germline-genotyped and -imputed SNPs within 1 Mb of the risk region with the expression levels of all genes within 1 Mb, up- and downstream, of the SNP in question (including *ADO* [MIM: 611392], *ARID5B* [MIM: 608538], *c10orf107*, *EGR2* [MIM: 129010], *JMJD1C* [MIM: 604503], *JMJD1C-AS1*, *REEP3* [MIM: 609348], *RTKN2* [MIM: 113705], and *ZNF365*) in normal breast, adjacent normal, and breast cancerous tissue from the following four cohorts. Normal breast I (NBI; n = 116) is comprised of women of European descent ascertained through multiple Norwegian hospitals. Gene expression data for the majority of women were derived from normal breast tissue in women who had not been affected with breast cancer; data for ten women were derived from normal tissue adjacent to a tumor. Genotyping was performed with the iCOGS SNP array, and gene expression levels were measured with the Agilent 44K array.[18,19] NBII (n = 93) is the European subset of the TCGA study, for whom expression data were available from normal tissue adjacent to a tumor. Germline genotype data from Affymetrix SNP 6 array, processed through Birdseed, were obtained from TCGA dbGAP data portal.[20] Gene expression levels were assayed by RNA sequencing, RSEM (RNaseq by Expectation-Maximization[21]) normalized per gene, as obtained from the TCGA consortium.[20] The data were $\log_2$ transformed, and unexpressed genes were excluded prior to eQTL analysis. Breast carcinomas I (BCI; n = 241) is a series of women of European ancestry diagnosed with breast cancer and recruited through multiple Norwegian hospitals. Genotypes were obtained with the iCOGS SNP array, and mRNA expression data were from the Agilent 44K array.[22] BCII (n = 765) is the TCGA breast cancer cohort; all non-European samples (as determined by clustering and PCA) were excluded from this analysis.[20] The genotyping platform was Affymetrix SNP 6, and gene expression data, for the breast tumors, was derived from RNA sequencing analysis, in a similar manner to NBII. There is no overlap between women recruited to each of these studies.

In addition, we examined all the genotyped or imputed SNPs used in the risk analysis for association with expression of nine genes (*ADO*, *ARID5B*, *c10orf107*, *EGR2*, *JMJD1C*, *NRBF2*, *REEP3*, *RTNK2*, and *ZNF365*, represented by 14 expression probes) in the 1 Mb region on either side of the fine-mapping interval, using data from normal tissue in patients from METABRIC. METABRIC (n = 135) comprises normal tissues adjacent to the tumors from breast cancer patients of genetically confirmed European ancestry from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) study.[23] The samples were assayed with the Illumina HT-12 v3 microarray. Matched germline SNP genotypes were available from the Affymetrix SNP 6.0 platform.

For all cohorts, the genotyping data were processed as follows: SNPs with call rates < 0.95 or minor allele frequencies < 0.05 were excluded, as were SNPs out of Hardy-Weinberg equilibrium with $p < 10^{-6}$. All samples with a call rate below 80% were excluded. Identity by state was computed with the R GenABEL package,[24] and closely related samples with IBS > 0.95 were removed. The SNP and sample filtration criteria were applied iteratively until all samples and SNPs met the stated thresholds. In total, 489 samples and 662,521 SNPs passed were kept in the analysis. Imputation was run on both the iCOGS and Affymetrix6 germline genotype data using the 1000 Genomes Project March 2012 v.3 release as the reference dataset.[25] A two-stage imputation procedure, using SHAPEIT to derive phased genotypes and IMPUTEv2 to perform the imputation on the phased data, has been found to significantly reduce the computational burden.[26]

Please cite this article in press as: Darabi et al., Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate *NRBF2* Expression, The American Journal of Human Genetics (2015), http://dx.doi.org/10.1016/j.ajhg.2015.05.002

The influence of germline genetic variations on gene expression was assessed using a linear regression model, as implemented in the R library eMAP. An additive effect was assumed by modeling copy number of the rare allele, i.e., 0, 1, or 2, for a given genotype. Only relationships in *cis*, i.e., where the SNP resided less than 1 Mb up or down from the center of the transcript, were investigated. Correction for multiple testing was performed using the false discovery rate (FDR) as implemented in the p.adjust function in R. Only FDR-adjusted, significant p values are reported in the Results section of the paper. Genotyping quality control and imputation for the METABRIC data are described in Guo et al.[27] For the METABRIC data, association between genotype and expression was tested by linear regression with FDR control as implemented in the MatrixEQTL package in R. We also conducted eQTL haplotype analysis in NBI and BCII which had been genotyped on iCOGS. The correlations with expression for the haplotypes residing in the four iCHAVs were estimated using the haplo.score function implemented in the R library haplo.stats, assuming a Gaussian distribution for the expression data.

Cell Lines

Normal breast epithelial cell lines MCF10A and Bre80 were grown in DMEM/F12 media supplemented with 5% horse serum, 10 μg/ml insulin, 0.5 μg/ml hydrocortisone, 20 ng/ml EGF, 100 ng/ml cholera toxin, and antibiotics. The MCF7 breast cancer cell line was grown in DMEM media supplemented with 10% fetal bovine serum, 10 μg/ml insulin, sodium pyruvate, and antibiotics. All cell lines were routinely tested for *Mycoplasma* and profiled with short tandem repeats to confirm their identity.

Chromatin Conformation Capture

Cross-linked DNA from the above cell lines was digested with EcoRI to generate chromatin conformation capture (3C) libraries as previously described.[17] 3C interactions were quantitated by real-time PCR, on at least two independent 3C libraries, using primers designed against the EcoRI restriction fragments across the 10q21.2 locus. Interactions were quantified in triplicate. The primers are listed in Table S2. Primer efficiencies were calculated using an artificial library of ligation products generated from two BAC clones (RP11-629KB and RP11-1021P21) that spanned the 10q21 locus and *NRBF2*. 3C interaction products were visualized by gel electrophoresis, gel purified, and sequenced to verify the 3C product.

In Silico Analysis

Genomic regions encompassing independent sets of correlated, highly trait-associated variants (iCHAVs) were examined for potential regulatory signals by overlaying epigenetic marks derived from ENCODE data using the UCSC Genome Browser. Tracks representing potential regulatory signals marked by DNaseI hypersensitivity, H3K4me1, H3K4me3, and H3K27ac histone modification, and transcription factor binding sites were obtained for the normal mammary cell types, human mammary epithelial cells (HMECs) and human mammary fibroblasts (HMFs), and two breast cancer cell lines, MCF7 and T47D. SNPs in each iCHAV were examined for potential functional consequences using HaploReg v.2[28] and RegulomeDB.[29]

Plasmid Generation

A 1,287-bp fragment containing the *ZNF365* promoter and a 1,163-bp fragment containing the *NRBF2* promoter were cloned into the pGL3 basic luciferase reporter. Using a sample heterozygous for rs2393886 as template, we then generated 1,875-bp PCR fragments containing part of iCHAV2 using PCR primers modified with BamH1 and Sal1 and cloned each haplotype into both the *ZNF365* and *NRBF2* promoter constructs. PCR primers are listed in Table S2.

Reporter Assays

Bre80, MCF10A, and MCF7 cells were transiently transfected with equimolar amounts of luciferase reporter constructs using *Renilla* luciferase as an internal control reporter. Luciferase activity was measured 24 hr after transfection using Dual-Glo Luciferase (Promega). To correct for any differences in transfection efficiency or cell lysate preparation, *Firefly* luciferase activity was normalized to *Renilla* luciferase, and the activity of each construct was measured relative to the promoter alone construct, which had a defined activity of 1. Association was assessed by log transforming the data and performing two-way ANOVA, followed by Dunnett's multiple comparisons test; for ease of interpretation, values were back transformed to the original scale for the graphs.

Results

We successfully genotyped 428 SNPs across the 560-kb fine-mapping region in 46,450 case subjects and 42,600 control subjects from 41 case-control studies in populations of European ancestry, and in 6,269 case subjects and 6,624 control subjects from 9 case-control studies of Asian ancestry.[5] We imputed genotypes for 3,409 SNPs in the interval in the European studies, at imputation $r^2 > 0.3$ and MAF $\geq$ 1%, using known genotypes in combination with data from the 1000 Genomes Project reference panel. Based on data from the European studies, 87 genotyped or imputed SNPs were convincingly associated with overall risk of breast cancer (p values 10^{-7} to 10^{-29}; Figure 1). The results for all SNPs in the interval with an association p value $< 1 \times 10^{-3}$ for overall breast cancer risk (382 SNPs) are presented in Table S3.

Among the genotyped SNPs, the strongest evidence of association with overall breast cancer risk among Europeans was for rs10995194 (OR [95% CI] = 0.86 [0.84–0.88], p = 3.77×10^{-29}). This SNP lies within intron 4 of *ZNF365* and is strongly correlated with the original GWAS hit rs10995190 ($r^2 = 0.99$). Analysis of the imputed SNPs identified two markers with slightly stronger associations: the most significantly associated marker was rs10995201 (OR = 0.85 [0.83–0.88], p = 1.05×10^{-29}), which is also strongly correlated with rs10995194 ($r^2 = 0.94$).

Multiple Independent Signals at 10q21.2

To determine whether there were additional independent signals of association at the locus, we included in a forward stepwise regression model 230 SNPs with MAF $\geq$ 2% that displayed evidence of association with overall breast cancer risk at p $< 10^{-4}$. The most parsimonious model included four independent SNPs that mark four iCHAVs (Figure 1; Table 1). These were as follows: iCHAV1, rs10995201 (OR = 0.85 [0.83–0.88], p = 1.05×10^{-29};

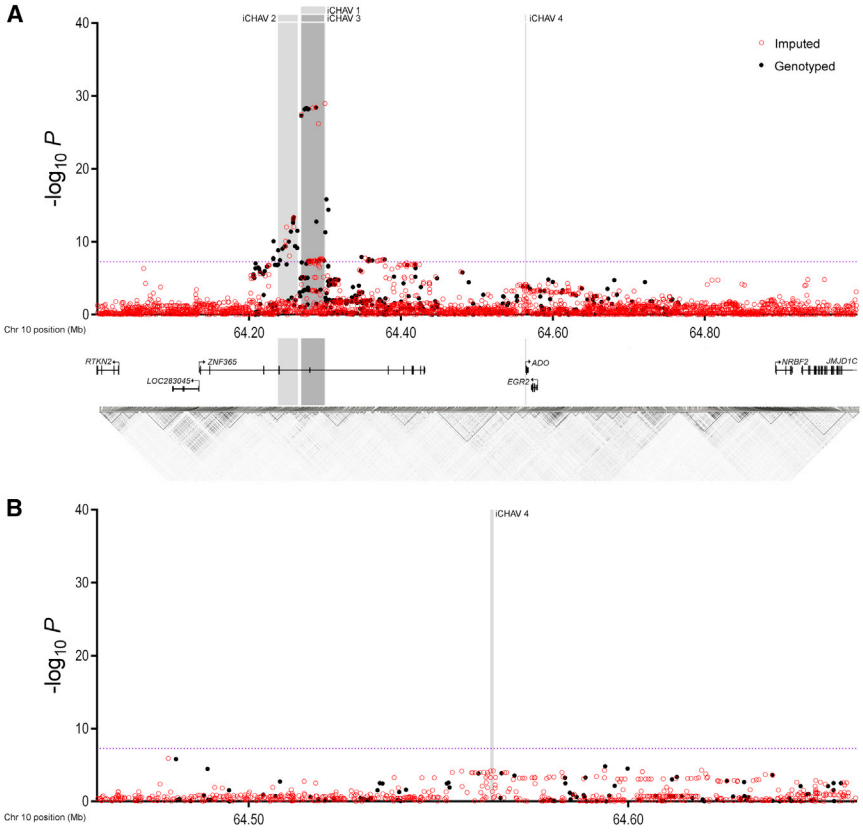

Figure 1. Association Results for Overall Breast Cancer Risk

Directly genotyped SNPs are shown as filled black circles, and imputed SNPs ($r^2 > 0.3$, MAF > 0.02) are shown as open red circles, plotted as the negative log of the p value against relative position across the locus.

(A) A schematic of the gene structures is shown. iCHAVs, encompassing all SNPs with a likelihood ratio of < 1:100 compared with the most significant SNP, are labeled and are shown as gray regions. The pattern of LD for all SNPs from the 1000 Genomes Project CEU population is shown as a plot of pairwise r^2 values using a greyscale, where white and black signify $r^2 = 0$ and 1, respectively. The dashed purple line represents genome-wide significance ($p < 5 \times 10^{-8}$).

(B) iCHAV4 is shown in more detail.

conditional p = 4.92×10^{-16}); iCHAV2, chr10: 64,258,684:D (OR = 0.93 [0.91–0.95], p = 4.24×10^{-14}; conditional p = 3.24×10^{-5}); iCHAV3, rs7922449 (OR = 1.06 [1.04–1.08], p = 1.68×10^{-7}; conditional p = 5.78×10^{-5}); iCHAV4, rs9971363 (OR = 0.94 [0.92–0.97], p = 6.54×10^{-5}; conditional p = 3.95×10^{-4}) (Figure 1B). These four SNPs were all imputed, with IMPUTE2 info-score above 0.85. The strongest pairwise LD r^2 value among the four markers was 0.12 (between rs10995201 and chr10: 64,258,684:D). Among the genotyped SNPs, the top SNP correlated with the lead iCHAV1 SNP was rs10995194 (OR = 0.86 [0.84–0.88], p = 3.77×10^{-29}, $r^2 = 0.94$), with the lead iCHAV2 SNP was rs2393886 (OR = 0.93 [0.91–0.95], p = 4.50×10^{-14}, $r^2 = 0.99$), with the lead iCHAV3 SNP was rs4746428 (OR = 1.06 [1.04–1.08], p = 3.24×10^{-8}, $r^2 = 0.68$), and with the lead iCHAV4 SNP was rs10995312 (OR = 0.95 [0.92–0.97], p = 1.38×10^{-4}, $r^2 = 0.97$).

In order to identify the candidate causal SNPs in each iCHAV for subsequent functional analysis, we calculated the likelihood ratio of each SNP relative to the best independent signal with which it was correlated ($r^2 > 0.6$), after adjusting for the lead SNP of preceding iCHAV(s). SNPs with a relative likelihood ratio of <1:100 compared with the most significant SNP for each iCHAV were excluded from consideration as being potentially causative.[16,30] Eleven SNPs had a relative likelihood ratio of >1:100 compared with the most significant SNP (rs10995201) and hence could not be excluded as causative for the lead signal—these SNPs were all strongly correlated with rs10995201 and span an interval of 31.2 kb (iCHAV1, Figure 1, Table S4). A 12th SNP, a single base insertion chr10: 64,291,099, was excluded at this threshold, but not strongly so (likelihood ratio ~1:600); all other SNPs could be clearly excluded (likelihood ratios < $1:10^{12}$). After adjustment for lead SNP of iCHAV1, we identified 16 strongly correlated SNPs and with a likelihood ratio >1:100 relative to the lead SNP in iCHAV2 (chr10: 64,258,684:D); these SNPs span an interval of 25 kb (iCHAV2, Figure 1, Table S5). SNP rs16917302 (OR = 0.96 [0.93–0.99], p = 1.03×10^{-2}), reported by Couch et al.,[31] lies within the region that iCHAV2 spans. However, rs16917302 is only weakly correlated with the lead SNP ($r^2 = 0.12$) in iCHAV2. After adjusting for the effects of lead SNPs of iCHAV1 and iCHAV2, 17 highly correlated SNPs had a relative likelihood >1:100 compared to rs7922449 (iCHAV3, Figure 1, Table S6) and span an interval of 29.1 kb. Finally, only one other SNP, rs7090365, had a relative likelihood of >1:100 with respect to rs9971363 the lead SNP of iCHAV4, after adjusting for the lead SNPs in iCHAVs 1–3 (Figure 1, Table S7). iCHAVs 1, 2, and 3 all lie within *ZNF365*. The SNPs in iCHAV1 and iCHAV3 span regions that physically overlap, whereas the SNPs in iCHAV2 lie telomeric to iCHAV1 and iCHAV3. iCHAV4 lies 5′ of *ADO*, which encodes 2-aminoethanethiol dioxygenase (Figure 1).

Association with Breast Cancer Subtypes

Based on data from European studies, 31 genotyped SNPs and 14 imputed SNPs were associated with risk of ER-positive breast cancer (p values 10^{-7} to 10^{-23}). The most strongly associated SNP for overall breast cancer (rs10995201) was also the most strongly associated SNP

Table 1. Four iCHAVs at 10q21.2 Associated with Breast Cancer Risk or Mammographic Density Phenotype in Europeans

iCHAV1				iCHAV2			iCHAV3			iCHAV4		
rs10995201[a]				chr10: 64,258,684:D[a]			rs7922449[a]			rs9971363[a]		
	p Value[b]	OR (95% CI)[b]		p Value[b]	OR (95% CI)[b]		p Value[b]	OR (95% CI)[b]		p Value[b]	OR (95% CI)[b]	
Overall	1.05×10^{-29}	0.85 (0.83–0.88)		4.24×10^{-14}	0.93 (0.91–0.95)		1.68×10^{-7}	1.06 (1.04–1.08)		6.54×10^{-5}	0.94 (0.92–0.97)	
ER^+	2.51×10^{-23}	0.85 (0.82–0.88)		8.01×10^{-11}	0.93 (0.91–0.95)		6.70×10^{-6}	1.06 (1.03–1.09)		1.26×10^{-5}	0.93 (0.90–0.96)	
ER^-	9.60×10^{-8}	0.87 (0.82–0.91)		0.0055	0.95 (0.91–0.98)		0.000268	1.08 (1.04–1.13)		0.776	0.99 (0.94–1.05)	
	p Value[c]	β (SE)[c]		p Value[c]	β (SE)[c]		p Value[c]	β (SE)[c]		p Value[c]	β (SE)[c]	
Density area	1.45×10^{-7}	−0.25 (0.05)		0.00555	−0.09 (0.03)		0.308	0.04 (0.04)		0.634	−0.02 (0.05)	
Percent density	1.32×10^{-5}	−0.15 (0.04)		0.00155	−0.08 (0.02)		0.634	0.01 (0.03)		0.629	−0.02 (0.04)	
	SNP[d]	p Value[d]	Info[e]	SNP[d]	p Value[d]	Info[e]	SNP[d]	p Value[d]	Info[e]	SNP[d]	p Value[d]	Info[e]
	rs10995181[f]	5.07×10^{-28}	1	rs4489633[f]	1.45×10^{-9}	0.96	rs1878253[f]	6.54×10^{-8}	1	rs9971363[a,g]	6.54×10^{-5}	0.97
	rs10995182[g]	3.42×10^{-28}	0.92	rs4282885[f]	8.51×10^{-10}	0.99	rs1914200[f]	1.03×10^{-7}	1	rs7090365[g]	6.59×10^{-5}	0.97
	rs10995187[f]	6.66×10^{-29}	1	rs10995173[f]	3.89×10^{-10}	1	rs10740081[g]	9.17×10^{-8}	0.99			
	rs4746419[f]	4.43×10^{-29}	1	rs10822012[g]	3.93×10^{-10}	0.99	rs10761639[g]	1.01×10^{-7}	0.98			
	rs34511355[f]	6.22×10^{-29}	1	rs10761637[g]	3.92×10^{-10}	0.99	rs7901318[g]	7.26×10^{-8}	0.98			
	rs10995189[f]	6.04×10^{-29}	1	rs12098307[g]	4.16×10^{-10}	0.99	rs2393894[g]	5.95×10^{-8}	0.97			
	rs10995190[f]	5.61×10^{-29}	1	rs10822013[f]	8.81×10^{-11}	1	rs7922449[a,g]	1.68×10^{-7}	0.86			
	rs10995191[f]	6.04×10^{-29}	1	rs10509168[f]	2.22×10^{-13}	1	rs10995196[g]	2.70×10^{-8}	0.98			
	rs11524313[g]	4.16×10^{-29}	0.97	rs10995176[g]	7.17×10^{-14}	0.99	rs4746428[f]	3.24×10^{-8}	1			
	rs10995193[g]	3.67×10^{-29}	0.99	c10_pos64258017[f]	1.03×10^{-13}	1	chr10: 64,293,571:D[g]	1.94×10^{-8}	0.88			
	rs10995194[f]	3.77×10^{-29}	1	rs2393886[f]	4.50×10^{-14}	1	rs9633558[g]	2.62×10^{-8}	0.99			
	rs10995201[a,g]	1.05×10^{-29}	0.95	chr10: 64,258,684:D[a,g]	4.24×10^{-14}	0.98	rs7915519[g]	2.63×10^{-8}	0.99			
				chr10: 64,258,692:D[g]	8.51×10^{-13}	0.86	rs6479823[g]	2.66×10^{-8}	0.99			
				rs12243471[f]	4.00×10^{-10}	0.99	rs1914182[g]	2.62×10^{-8}	0.99			
				rs12245332[f]	3.95×10^{-10}	1	rs10822017[g]	2.07×10^{-8}	0.99			
				rs2393887[f]	6.94×10^{-10}	1	rs7901573[g]	2.93×10^{-8}	0.99			
				rs4746409[f]	3.06×10^{-12}	1	rs12258134[g]	2.94×10^{-8}	0.99			
							rs1949356[f]	3.61×10^{-8}	1			

[a]Lead SNP for each iCHAV.
[b]p value, odds ratio (OR), and 95% confidence interval (CI) for association with overall breast cancer risk, estrogen-positive (ER^+), and estrogen-negative (ER^-) disease.
[c]p value, beta, and standard error (SE) for association with percentage density, dense area.
[d]SNPs correlated ($r^2 > 0.6$) with a likelihood ratio of >1:100 with respect to overall risk association relative to lead SNP of each iCHAV and corresponding p value for association with overall breast cancer risk.
[e]IMPUTE2 info score.
[f]Genotyped marker.
[g]Imputed marker.

Please cite this article in press as: Darabi et al., Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate *NRBF2* Expression, The American Journal of Human Genetics (2015), http://dx.doi.org/10.1016/j.ajhg.2015.05.002

for ER-positive disease (OR = 0.85 [0.82–0.88], p = 2.51 × 10^{-23}) and had a similar association for ER-negative disease (OR = 0.87 [0.82–0.91], p = 9.60 × 10^{-08}, $p_{heterogeneity}$ = 0.34) (Table 1, Figures S1 and S2). The most strongly associated, well-imputed SNP (IMPUTE2 info-score > 0.9) for ER-negative disease was rs10995182 (OR = 0.85 [0.8–0.9], p = 6.68 × 10^{-8}), another SNP within iCHAV1 strongly correlated with rs10995201. Thus, the results indicate that susceptibility SNPs in iCHAV1 confer similar relative risks for ER-positive and ER-negative disease. Similarly, the lead SNPs in iCHAVs 2 and 3 showed similar associations for ER-positive and ER-negative disease (OR_{ER+} = 0.96 [0.94–0.98], p_{ER+} = 8.77 × 10^{-04}, OR_{ER-} = 0.98 [0.94–1.02], p_{ER-} = 0.309, $p_{heterogeneity}$ = 0.44 for chr10: 64,258,684:D adjusted for iCHAV1, OR_{ER+} = 1.04 [1.02–1.07], p_{ER+} = 1.54 × 10^{-03}, OR_{ER-} = 1.07 [1.02–1.11], p_{ER-} = 4.36 × 10^{-03}, $p_{heterogeneity}$ = 0.44 for rs7922449 adjusted for iCHAVs 1 and 2). However, the lead SNP in iCHAV4 showed no association with ER-negative disease (OR_{ER+} = 0.93 [0.90–0.97], p_{ER+} = 7.72 × 10^{-05}, OR_{ER-} = 1.00 [0.94–1.05], p_{ER-} = 0.862, $p_{heterogeneity}$ = 0.016 for rs9971363 adjusted for iCHAVs 1, 2, and 3).

To determine whether there were additional subtype-specific signals of association, we included all SNPs displaying evidence for association with ER-positive disease (222 SNPs, p < 10^{-4} and MAF ≥ 2%) and ER-negative disease (19 SNPs, p < 10^{-4} and MAF ≥ 2%) in separate forward stepwise regression models. For ER-positive disease, two iCHAVs were identified: iCHAV1-ER+ rs10995201 (OR = 0.85 [0.83–0.88], p = 2.51 × 10^{-23}; conditional p = 1.65 × 10^{-18}) in iCHAV1 and iCHAV2-ER+ chr10: 64,258,684:D (OR = 0.93 [0.91–0.95], p = 8.01 × 10^{-11}; conditional p = 2.59 × 10^{-05}) in iCHAV2.

Breast Cancer Risk Associations in Asian Studies

The top associated marker among Asians was the genotyped SNP rs7914770 (OR = 0.93 [0.89–0.98], p = 0.006). This SNP, which lies between iCHAV1 and iCHAV4, as defined in analysis of the European population, was not associated with overall breast cancer risk among Europeans (OR = 1.01 [0.98–1.03], p = 0.534; Table S8). The most strongly associated genotyped SNP in Europeans, rs10995194 in iCHAV1, showed a borderline association with breast cancer risk in Asians, but in the opposite direction (Asians: OR = 1.18 [1.00–1.39], p = 0.04; Europeans: OR = 0.86 [0.84–0.88], p = 3.77 × 10^{-29}). The minor (C) allele of rs10995194 (MAF = 0.15) and the other iCHAV1 SNPs were much rarer in Asians (MAF = 0.02), but the ORs estimates nevertheless differed significantly (p = 10^{-4}). In iCHAV2, the most strongly associated genotyped SNP in Europeans, rs2393886 (OR = 0.93, MAF = 0.47 for the A allele), was also associated with risk in Asians; however, in contrast to iCHAV1, the effects were in the same direction and of comparable magnitude (OR = 0.95 [0.90–1.00], p = 0.04, allele frequency 0.51 for the A allele in Asians). In this iCHAV the most strongly associated SNP in Asians was rs4746409 (OR = 1.06 [1.00–1.11], p = 0.03), but this is strongly correlated with the lead SNP rs2393886 (r^2 = 0.88). SNP rs10822013 within iCHAV2, recently identified by Cai et al. in an Asian GWAS,[7] was more weakly associated with risk in Asians in our data (OR = 0.96 [0.90–1.10], p = 0.07). None of the genotyped markers in iCHAV3 or iCHAV4 that were significant in Europeans showed association with overall breast cancer risk in the Asian population.

Haplotype Analysis

We conducted haplotype analysis using the most probable genotype for the imputed lead SNPs in each of the iCHAVs. We grouped haplotypes with a frequency < 0.01 into one (rare) group. The haplo.score procedure estimated 11 haplotypes with non-zero frequencies for all three phenotypes (overall, ER^+, and ER^- risk) (Table S9). The most significant association was observed for haplotype H6, carrying the rare allele of the lead SNPs of iCHAV1 and iCHAV2: OR = 0.86 (0.82–0.90), p = 8.52 × 10^{-10}. Consistent with the regression analyses, the three haplotypes (H6, H14, H5) carrying the rare allele of the iCHAV1 SNP rs10995201 were all associated with a similar, reduced breast cancer risk. Three other haplotypes, all carrying the rare allele of the iCHAV2 SNP, chr10: 64,258,684:D, were also associated with a reduced risk, relative to the baseline haplotype, consistent with an independent effect of iCHAV2.

Association with Mammographic Density Phenotypes

Multiple linear regression was used to investigate the association between mammographic density phenotypes (PD, percent density; DA, dense area; nDA, non-dense area) and genotypes after adjustment for other covariates. Among the three density phenotypes, the strongest associations were observed with DA. The strongest association with DA and PD was seen with the lead SNP of iCHAV1, rs10995201 (β (SE) = −0.25 [0.05], p = 1.45 × 10^{-07} for DA, β (SE) = −0.15 [0.04], p = 1.32 × 10^{-05} for PD; Tables S10, S11; Figures S3 and S4). After adjusting for the lead SNP in iCHAV1, no SNPs in iCHAV2 were associated with any of the density phenotypes at p < 0.01 (Table S12); similarly, none of the iCHAV3 or iCHAV4 SNPs were associated with density phenotypes (Tables S13 and S14). The strongest association with nDA was seen with imputed marker rs224303 about 33 kb away from iCHAV4 (Figure S5), which was not associated with overall risk of breast cancer (p = 0.745).

To assess the extent to which the observed association with breast cancer risk might be mediated through a mammographic density phenotype, we estimated the association with breast cancer risk before and after adjustment for PD, DA, or nDA in a pooled analysis of the 2,379 breast cancer case subjects and 4,507 control subjects for whom density measurements were available. The association between lead iCHAV1 SNP rs10995201 (i.e., the only iCHAV for which an association with mammographic density had been identified) and breast cancer in this subset (OR = 0.87

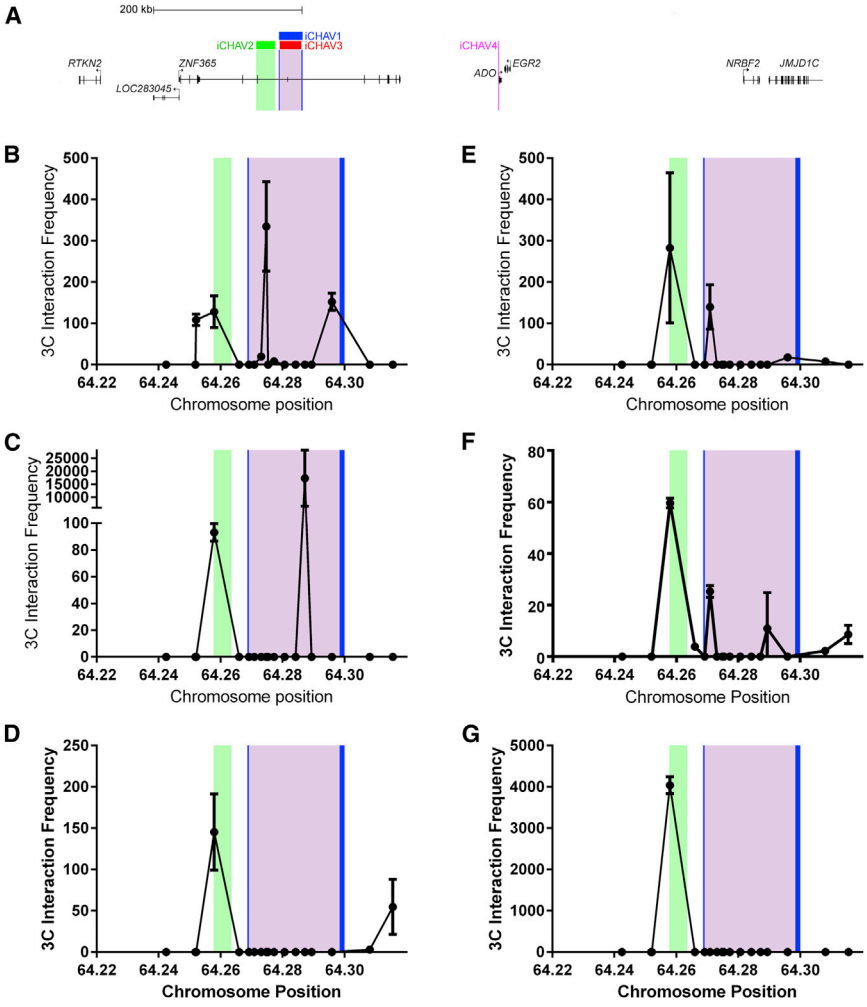

Figure 2. Chromatin Interactions with the *ZNF365* and *NRBF2* Promoters at the 10q21.2 Locus

(A) 10q21.2 locus showing the distances from iCHAVs 1–4 and the nearest genes.

(B–G) Chromatin interaction frequencies were plotted at the corresponding chromosomal position for MCF7 (B and E), MCF10A (C and F), and Bre80 (D and G) for the *ZNF365* and *NRBF2* promoters, respectively. iCHAV2 is marked in green, iCHAV1 (which physically overlaps iCHAV3) in blue, and iCHAV3 in red. Representative graphs are shown (N = 3) and error bars denote SD.

[0.77–0.98], p = 0.02) was similar to that in the complete BCAC set and was only slightly attenuated after adjustment for PD (OR = 0.88 [0.78–1.00], p = 0.05) or DA (OR = 0.89 [0.79–1.01], p = 0.06) (Table S15).

eQTL Analysis

We analyzed 2,238, 5,122, 2,250, 5,211, and 3,814 SNPs in NBI, NBII, BCI, BCII, and METABRIC, respectively, for association with the expression levels of all genes within 1 Mb, up- and downstream, of the SNP in question. Significant eQTL associations were observed for both normal breast and tumors. Multiple SNPs within the fine mapping region associated with expression of *c10orf107* in the NBII cohort (strongest association, chr10: 63,427,159:D, which is not a candidate causal risk SNP; FDR-adjusted $p = 6.1 \times 10^{-5}$ and $r^2 = 0.325$). For the BCI cohort, multiple eQTLs were found for both *c10orf107* and *RTKN2* (strongest associations chr10: 63,427,159:D, FDR-adjusted $p = 4.3 \times 10^{-6}$ and $r^2 = 0.15$, and rs870988, FDR-adjusted $p = 1.4 \times 10^{-5}$ and $r^2 = 0.13$, respectively). In BCII we observed multiple eQTLs for three different genes, *c10orf107*, *EGR2*, and *ADO* (rs12781009, FDR-adjusted $p = 5.6 \times 10^{-9}$ and $r^2 = 0.066$; rs34632941, FDR-adjusted p = 0.02 and $r^2 = 0.02$; and rs224045, FDR-adjusted p = 0.02 and $r^2 = 0.02$, respectively). No significant associations with expression were found for NBI or METABRIC, and there were no significant associations between the putative causal SNPs in iCHAVs 1–4 and expression of any of the genes analyzed in the region in any study (Tables S16 and S17). Significant associations (without correction for multiple testing) were found between haplotypes of iCHAV1 and *NRBF2* in NBI and BCI (p = 0.005 and p = 0.042, respectively) and between haplotypes of iCHAV2 and *NRBF2* in NBI (p = 0.011). In addition, we found associations between haplotypes of iCHAV1 and *REEP3*, *EGR2*, and *RTKN2* (p = 0.011–0.049), haplotypes of iCHAV2 and *RTKN2* (p = 0.012), haplotypes of iCHAV3 and *REEP3*, *JMJD1C*, *ARID5B*, *ADO*, *RTKN2*, and *TMEM26* (p = 0.001–0.036), and haplotypes of iCHAV4 and *ARID5B* (p = 0.012) in either NBI or BCI, but not both.

Chromosome Conformation Capture Analyses Identify *ZNF365* and *NRBF2* as Target Genes

We performed 3C experiments to determine whether there were any chromatin interactions between the *ZNF365* and *NRBF2* promoters and iCHAVs 1, 2, and 3. We identified significant interactions between iCHAV2 at the 10q21.2 locus and both promoter regions in the breast cancer cell line, MCF7, and in two normal breast cell lines, MCF10A and Bre80 (Figures 2 and S6). No reproducible interactions were detected between iCHAVs 1 or 3 and *ZNF365* and *NRBF2* in the cell lines analyzed. Although we did observe some possible interactions in MCF7 and MCF10A between different parts of iCHAV 1/3 and the promoters of both *ZNF365* and *NRBF2*, they were not reproducible (Figures 2 and S6).

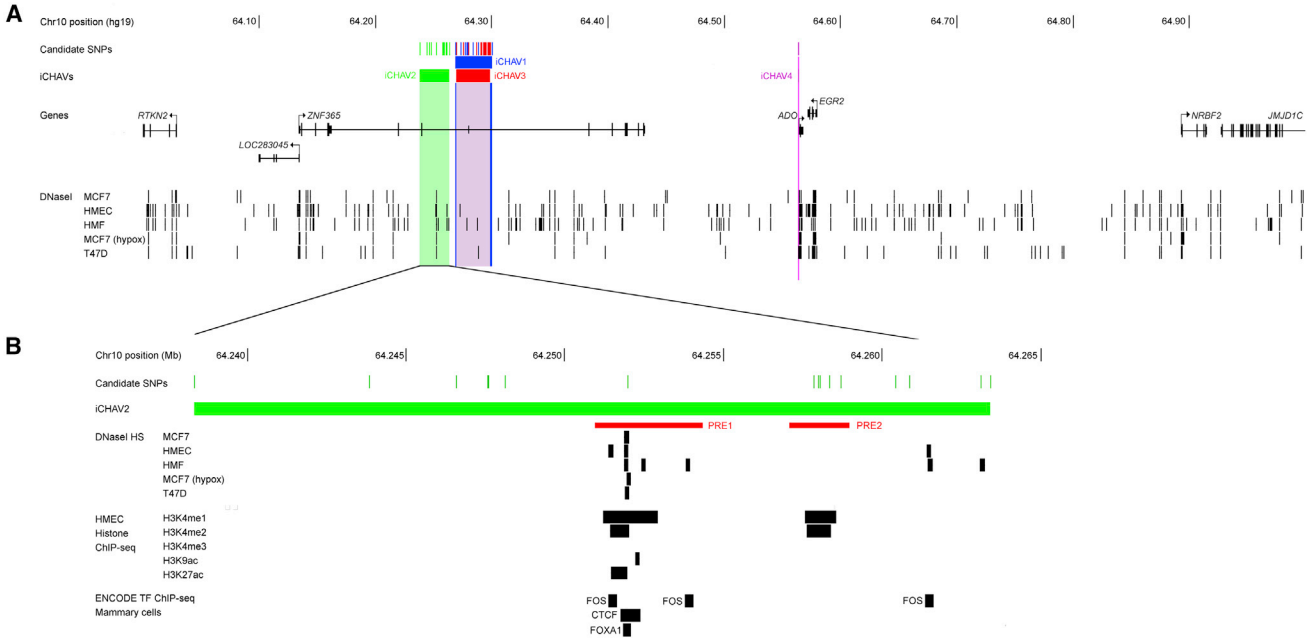

Figure 3. Chromatin Marks in Breast Cells in iCHAVs at 10q21.2
The region encompassing 1 Mb at 10q21.2 is shown in (A). Candidate causal SNPs lying within iCHAVs 1–4 are shown as tick marks in matching colors to iCHAVs. DNaseI hypersensitive sites found in mammary cell types from ENCODE are depicted under the gene schematics. iCHAV2 is shown in (B). ENCODE tracks are shown for mammary DNaseI HS, histone modification ChIP-seq, and transcription factor ChIP-seq. The cloned PRE2 region for the iCHAV2 reporter construct is marked.

Identification of a Putative Regulatory Region within iCHAV2 Regulating both *ZNF365* and *NRBF2*

We used available ENCODE ChIP-seq data to identify putative regulatory elements (PREs) within iCHAV2 (Figure 3A). We identified two PREs (PRE1 and PRE2), as defined by DNaseI hypersensitivity sites (indicative of regions of open chromatin), in several normal and cancer breast cell lines, and H3K4me1 and H3K4me2 histone modifications in HMECs (Figure 3B). PRE1 lies in a complex repetitive region and we were unable to clone and analyze this region. However, 6 (chr10: 64,258,684:D, rs2393886, rs10995176, c10_pos64258017, rs10509168, and chr10: 64,258,692:D) of the 17 candidate causal SNPs lie within PRE2 (Figure 3B). We examined the regulatory capability of PRE2, combined with the effect of the protective haplotype, using luciferase constructs containing the *ZNF365* or *NRBF2* promoters. Inclusion of the reference haplotype of iCHAV2 did not significantly alter the effect of the *NRBF2* promoter in Bre80 cells. However, the construct containing the protective haplotype acted as a silencer relative to the iCHAV2 reference allele ($p = 0.003$; Figures 4A and S7). A similar trend was seen in MCF7 cells ($p = 0.029$; Figure 4C). Although constructs containing the *ZNF365* promoter showed a trend reducing its activity with the reference haplotype of iCHAV2, this was significant only in MCF7 cells (Figures 4B and 4D). Similarly, inclusion of the protective haplotype reduced the *ZNF365* promoter activity but had no effect relative to the iCHAV2 reference haplotype in either Bre80 or MCF7 cells (Figures 4B and 4D).

Discussion

Our large combined dataset provides clear confirmation of a susceptibility locus for breast cancer at 10q21.2 for breast cancer as originally reported by Turnbull et al.[4] and for mammographic density as reported by Lindström et al.[9] Multiple regression and haplotype analyses showed clear evidence of at least two, and potentially four, independent susceptibility loci in this region in Europeans. The most strongly associated SNP, rs10995201 in iCHAV1, showed clear evidence of association with both ER-positive and ER-negative disease, with the ORs being similar, and was also the SNP most strongly associated with mammographic density (DA, and hence PD). iCHAV2 (lead SNP, chr10: 64,258,684) and iCHAV3 (lead SNP, rs7922449) also appeared to be associated with both ER-positive and ER-negative disease. Evidence of a fourth iCHAV, associated only with ER-positive breast cancer, was weaker. In contrast to the results in Europeans, there was less evidence of association between SNPs at this locus and breast cancer risk in Asian women. The top associated genotyped SNP among Asians was rs7914770, which was not associated with overall breast cancer risk among Europeans. Furthermore, the lead genotyped SNP in Europeans, rs10995194, showed borderline evidence of association with breast cancer in Asians in the opposite direction, such that the effect sizes were clearly different. There are several possible explanations for the difference in the effect of iCHAV1 by ethnicity. The difference might reflect the differential effect of another SNP, or SNPs, in the

Please cite this article in press as: Darabi et al., Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate *NRBF2* Expression, The American Journal of Human Genetics (2015), http://dx.doi.org/10.1016/j.ajhg.2015.05.002

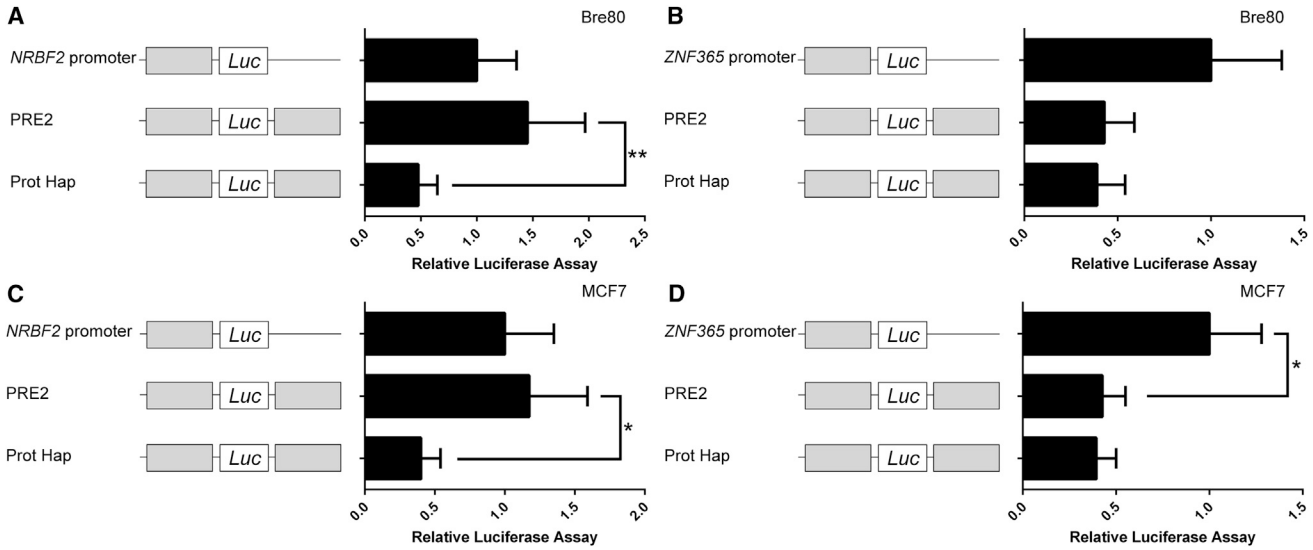

Figure 4. Protective Haplotype of iCHAV2 Silences *NRBF2* Promoter Activity
Both haplotypes of the iCHAV2 PRE2 were cloned upstream of *NRBF2*- and *ZNF365*-promoter-driven luciferase reporters. Cells were transiently transfected with the common (PRE2) and the protective (Prot Hap) haplotype constructs and assayed for luciferase activity 24 hr later. Results for the *NRBF2* and *ZNF365* promoters in Bre80 cells are shown in (A) and (B) and results for *NRBF2* and *ZNF365* promoters in MCF7 cells are shown in (C) and (D). Error bars denote 95% confidence intervals from three independent experiments performed in triplicate. p values were determined by two-way ANOVA followed by Dunnett's multiple comparisons test (**$p < 0.01$, *$p < 0.05$) on log transformed data; for ease of interpretation, back-transformed data have been graphed.

10q21.2 region. Alternatively, it might reflect *trans*-interactions with SNPs elsewhere in the genome. Larger studies in Asian populations, both for breast cancer risk and mammographic density, might help to resolve this paradox. The SNPs in iCHAV2, however, showed effects that were consistent between the two populations.

The *A* allele of rs10995190 in iCHAV1 is associated with decreased ER-positive and ER-negative breast cancer risk, as well as with lower percent mammographic density and reduced dense area. These results are consistent with the hypothesis that the same causal SNP confers susceptibility to both mammographic density and breast cancer risk. However, adjustment for percent density or dense area caused only a minor attenuation of the association between breast cancer risk and the lead iCHAV1 SNP. The implications of this finding are not clear, but it is likely that percent and absolute mammographic density as captured by a two-dimensional mammogram, despite being strong predictors of breast cancer risk, are imperfect measures of the underlying mechanism that drives breast cancer risk. Alternatively, it might be that different SNPs in this iCHAV are acting independently on mammographic density and breast cancer risk, and through different mechanisms. We did not observe associations with mammographic density for breast cancer-associated SNPs in the other iCHAVs, but the smaller sample size, and hence lower statistical power, for the mammographic density analysis mean that more subtle effects in other iCHAVs would not have been detectable.

As we have reported for the breast cancer loci at 11q13[17] and 2q35,[32] we did not find any evidence for single SNP eQTLs for any of the putatively causal SNPs in iCHAVs 1–4 in normal or cancerous breast samples. This might be because the power of the eQTL studies (n = 93–765) is limited, or because eQTLs are context dependent and might be expressed only in certain cell types or in response to certain stimuli. We have, however, identified a putative regulatory element in iCHAV2 at the 10q21.2 locus that interacts with both the *NRBF2* and *ZNF365* promoters in both normal breast and breast epithelial tumor cells. It is interesting that we detected haplotype associations with iCHAV1 and iCHAV2 and *NRBF2* expression, but the significance of these is difficult to interpret given their modest p values and the fact that these iCHAVs are far too big to be cloned in their entirety for luciferase assays.

We prioritized the 3C analysis to look for interactions between iCHAVs 1, 2, and 3 and the promoters of *ZNF365* and *NRBF2* because of their function. *NRBF2* encodes nuclear receptor binding protein 2, an interaction partner of the peroxisome proliferator-activated receptor alpha (PPARα), and exhibits a gene activation function in mammalian cells.[33] Furthermore, NRFB2 is thought to have a role in cell survival in neural progenitor cells[34] and to suppress autophagy,[35] a process that needs to be tightly controlled in breast cells during normal development, tissue differentiation, and response to stress.[36] To our knowledge, NRBF2 has not previously been implicated in breast cancer tumorigenesis. Our results suggest that the haplotype in iCHAV2 associated with reduced risk of breast cancer is associated with silencing of the *NRBF2* promoter. *ZNF365* encodes the zinc finger protein 365, which plays a critical role in stabilizing fragile sites within the genome and telomeres[37] and maintaining genome stability,[38] and is therefore a good candidate for a breast cancer

susceptibility gene. Although we did not observe a differential effect of the protective haplotype on the *ZNF365* promoter in luciferase assays, the protective SNPs might act through differential looping as we have previously observed at the 2q35 breast cancer risk locus.[32] Alternatively, other candidate causal SNPs in PRE1 or elsewhere in iCHAV2 might have a differential effect on transactivation of the *ZNF365* promoter. We did not find any convincing evidence for interactions between iCHAV1/3 and the promoters of *ZNF365* or *NRBF2* in MCF7, MCF10A, or Bre80 cells. It is possible that other genes, such as *ADO*, which encodes 2-aminoethanethiol dioxygenase, or *EGR2* encoding early growth response-2, are the targets of these iCHAVs, or that their interactions with *ZNF365* or *NRBF2* are manifest in different cell types, such as of the immune system, or only in response to specific stimuli.

In conclusion, we have found evidence for four sets of correlated genetic variants (iCHAVs) at 10q21.2 independently associated with breast cancer risk, one of which is also associated with mammographic density. In one of these iCHAVs, we have identified candidate causal SNPs that affect expression of *NRBF2*, which lies more than 600 kb away, suggesting that expression of NRFB2 might play a role in transformation or progression of transformed breast cells.

Supplemental Data

Supplemental Data include Supplemental Acknowledgments, 7 figures, and 17 tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2015.05.002.

Received: February 8, 2015
Accepted: May 1, 2015
Published: June 11, 2015

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://browser.1000genomes.org
BLAST, http://blast.ncbi.nlm.nih.gov/Blast.cgi
eMAP, http://www.bios.unc.edu/~weisun/software.htm
HaploReg, http://www.broadinstitute.org/mammals/haploreg/haploreg.php
NCBI, http://www.ncbi.nlm.nih.gov/
OMIM, http://www.omim.org/
RegulomeDB, http://RegulomeDB.org/
UCSC Genome Browser, http://genome.ucsc.edu

References

1. Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., and Parkin, D.M. (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int. J. Cancer *127*, 2893–2917.
2. Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer *136*, E359–E386.
3. Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., and Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. N. Engl. J. Med. *343*, 78–85.
4. Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., Seal, S., Ghoussaini, M., Hines, S., Healey, C.S., et al.; Breast Cancer Susceptibility Collaboration (UK) (2010). Genome-wide association study identifies five new breast cancer susceptibility loci. Nat. Genet. *42*, 504–507.
5. Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K., et al.; Breast and Ovarian Cancer Susceptibility Collaboration; Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON); kConFab Investigators; Australian Ovarian Cancer Study Group; GENICA (Gene Environment Interaction and Breast Cancer in Germany) Network (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. *45*, 353–361, e1–e2.
6. Zheng, W., Zhang, B., Cai, Q., Sung, H., Michailidou, K., Shi, J., Choi, J.Y., Long, J., Dennis, J., Humphreys, M.K., et al. (2013). Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. Hum. Mol. Genet. *22*, 2539–2550.
7. Cai, Q., Long, J., Lu, W., Qu, S., Wen, W., Kang, D., Lee, J.Y., Chen, K., Shen, H., Shen, C.Y., et al. (2011). Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. Hum. Mol. Genet. *20*, 4991–4999.
8. Antoniou, A.C., Kuchenbaecker, K.B., Soucy, P., Beesley, J., Chen, X., McGuffog, L., Lee, A., Barrowdale, D., Healey, S., Sinilnikova, O.M., et al.; CIMBA, SWE-BRCA; HEBON; EMBRACE; GEMO Collaborators Study; kConFab Investigators (2012). Common variants at 12p11, 12q24, 9p21, 9q31.2 and in ZNF365 are associated with breast cancer risk for BRCA1 and/or BRCA2 mutation carriers. Breast Cancer Res. *14*, R33.
9. Lindström, S., Vachon, C.M., Li, J., Varghese, J., Thompson, D., Warren, R., Brown, J., Leyland, J., Audley, T., Wareham, N.J., et al. (2011). Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. Nat. Genet. *43*, 185–187.
10. Eriksson, N., Benton, G.M., Do, C.B., Kiefer, A.K., Mountain, J.L., Hinds, D.A., Francke, U., and Tung, J.Y. (2012). Genetic variants associated with breast size also influence breast cancer risk. BMC Med. Genet. *13*, 53.
11. Varghese, J.S., Thompson, D.J., Michailidou, K., Lindström, S., Turnbull, C., Brown, J., Leyland, J., Warren, R.M., Luben, R.N., Loos, R.J., et al.; MODE Consortium (2012). Mammographic breast density and breast cancer: evidence of a shared genetic basis. Cancer Res. *72*, 1478–1484.
12. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
13. Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands of genomes. G3 (Bethesda) *1*, 457–470.

14. Vachon, C.M., Scott, C.G., Fasching, P.A., Hall, P., Tamimi, R.M., Li, J., Stone, J., Apicella, C., Odefrey, F., Gierach, G.L., et al. (2012). Common breast cancer susceptibility variants in LSP1 and RAD51L1 are associated with mammographic density measures that predict breast cancer risk. Cancer Epidemiol. Biomarkers Prev. *21*, 1156–1166.
15. Byng, J.W., Boyd, N.F., Fishell, E., Jong, R.A., and Yaffe, M.J. (1994). The quantitative analysis of mammographic densities. Phys. Med. Biol. *39*, 1629–1638.
16. Ursin, G., Ma, H., Wu, A.H., Bernstein, L., Salane, M., Parisky, Y.R., Astrahan, M., Siozon, C.C., and Pike, M.C. (2003). Mammographic density and breast cancer in three ethnic groups. Cancer Epidemiol. Biomarkers Prev. *12*, 332–338.
17. French, J.D., Ghoussaini, M., Edwards, S.L., Meyer, K.B., Michailidou, K., Ahmed, S., Khan, S., Maranian, M.J., O'Reilly, M., Hillman, K.M., et al.; GENICA Network; kConFab Investigators (2013). Functional variants at the 11q13 risk locus for breast cancer regulate cyclin D1 expression through long-range enhancers. Am. J. Hum. Genet. *92*, 489–503.
18. Quigley, D.A., Fiorito, E., Nord, S., Van Loo, P., Alnæs, G.G., Fleischer, T., Tost, J., Moen Vollan, H.K., Tramm, T., Overgaard, J., et al. (2014). The 5p12 breast cancer susceptibility locus affects MRPS30 expression in estrogen-receptor positive tumors. Mol. Oncol. *8*, 273–284.
19. Haakensen, V.D., Lingjaerde, O.C., Lüders, T., Riis, M., Prat, A., Troester, M.A., Holmen, M.M., Frantzen, J.O., Romundstad, L., Navjord, D., et al. (2011). Gene expression profiles of breast biopsies from healthy women identify a group with claudin-low features. BMC Med. Genomics *4*, 77.
20. Cancer Genome Atlas, N.; Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.
21. Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A., and Dewey, C.N. (2010). RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics *26*, 493–500.
22. Naume, B., Zhao, X., Synnestvedt, M., Borgen, E., Russnes, H.G., Lingjaerde, O.C., Strømberg, M., Wiedswang, G., Kvalheim, G., Kåresen, R., et al. (2007). Presence of bone marrow micrometastasis is associated with different recurrence risk within molecular subtypes of breast cancer. Mol. Oncol. *1*, 160–171.
23. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al.; METABRIC Group (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.
24. Aulchenko, Y.S., Ripke, S., Isaacs, A., and van Duijn, C.M. (2007). GenABEL: an R library for genome-wide association analysis. Bioinformatics *23*, 1294–1296.
25. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
26. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. *44*, 955–959.
27. Guo, Q., Schmidt, M.K., Kraft, P., Canisius, S., Chen, C., Khan, S., Tyrer, J., Bolla, M.K., Wang, Q., Dennis, J., et al.; kConFab Investigators (2015). Identification of novel genetic markers of breast cancer survival. J. Natl. Cancer Inst. *107*, djv081.
28. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. *40*, D930–D934.
29. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. *22*, 1790–1797.
30. Spencer, A.V., Cox, A., and Walters, K. (2014). Comparing the efficacy of SNP filtering methods for identifying a single causal SNP in a known association region. Ann. Hum. Genet. *78*, 50–61.
31. Couch, F.J., Gaudet, M.M., Antoniou, A.C., Ramus, S.J., Kuchenbaecker, K.B., Soucy, P., Beesley, J., Chen, X., Wang, X., Kirchhoff, T., et al.; OCGN; SWE-BRCA; HEBON; EMBRACE; GEMO Study Collaborators; kConFab investigators; Consortium of Investigators of Modifiers of BRCA1/2 (2012). Common variants at the 19p13.1 and ZNF365 loci are associated with ER subtypes of breast cancer and ovarian cancer risk in BRCA1 and BRCA2 mutation carriers. Cancer Epidemiol. Biomarkers Prev. *21*, 645–657.
32. Ghoussaini, M., Edwards, S.L., Michailidou, K., Nord, S., Cowper-Sal Lari, R., Desai, K., Kar, S., Hillman, K.M., Kaufmann, S., Glubb, D.M., et al.; Australian Ovarian Cancer Management Group; Australian Ovarian Cancer Management Group (2014). Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. Nat. Commun. *4*, 4999.
33. Yasumo, H., Masuda, N., Furusawa, T., Tsukamoto, T., Sadano, H., and Osumi, T. (2000). Nuclear receptor binding factor-2 (NRBF-2), a possible gene activator protein interacting with nuclear hormone receptors. Biochim. Biophys. Acta *1490*, 189–197.
34. Larsson, J., Forsberg, M., Brännvall, K., Zhang, X.Q., Enarsson, M., Hedborg, F., and Forsberg-Nilsson, K. (2008). Nuclear receptor binding protein 2 is induced during neural progenitor differentiation and affects cell survival. Mol. Cell. Neurosci. *39*, 32–39.
35. Zhong, Y., Morris, D.H., Jin, L., Patel, M.S., Karunakaran, S.K., Fu, Y.J., Matuszak, E.A., Weiss, H.L., Chait, B.T., and Wang, Q.J. (2014). Nrbf2 protein suppresses autophagy by modulating Atg14L protein-containing Beclin 1-Vps34 complex architecture and reducing intracellular phosphatidylinositol-3 phosphate levels. J. Biol. Chem. *289*, 26021–26037.
36. Clarke, R., Cook, K.L., Hu, R., Facey, C.O., Tavassoly, I., Schwartz, J.L., Baumann, W.T., Tyson, J.J., Xuan, J., Wang, Y., et al. (2012). Endoplasmic reticulum stress, the unfolded protein response, autophagy, and the integrated regulation of breast cancer cell fate. Cancer Res. *72*, 1321–1331.
37. Zhang, Y., Park, E., Kim, C.S., and Paik, J.H. (2013). ZNF365 promotes stalled replication forks recovery to maintain genome stability. Cell Cycle *12*, 2817–2828.
38. Zhang, Y., Shin, S.J., Liu, D., Ivanova, E., Foerster, F., Ying, H., Zheng, H., Xiao, Y., Chen, Z., Protopopov, A., et al. (2013). ZNF365 promotes stability of fragile sites and telomeres. Cancer Discov *3*, 798–811.