Functional Variants at the 11q13 Risk Locus for Breast Cancer Regulate Cyclin D1 Expression through Long-Range Enhancers

Juliet D. French,[1,131] Maya Ghoussaini,[2,131] Stacey L. Edwards,[1,131] Kerstin B. Meyer,[3,131] Kyriaki Michailidou,[4] Shahana Ahmed,[2] Sofia Khan,[5] Mel J. Maranian,[2] Martin O'Reilly,[3] Kristine M. Hillman,[1] Joshua A. Betts,[1] Thomas Carroll,[3] Peter J. Bailey,[1] Ed Dicks,[2] Jonathan Beesley,[6] Jonathan Tyrer,[2] Ana-Teresa Maia,[3] Andrew Beck,[7] Nicholas W. Knoblauch,[7] Constance Chen,[8] Peter Kraft,[8,9] Daniel Barnes,[4] Anna González-Neira,[10] M. Rosario Alonso,[10] Daniel Herrero,[10] Daniel C. Tessier,[11] Daniel Vincent,[11] Francois Bacot,[11] Craig Luccarini,[2] Caroline Baynes,[2] Don Conroy,[2] Joe Dennis,[4] Manjeet K. Bolla,[4] Qin Wang,[4] John L. Hopper,[12] Melissa C. Southey,[13] Marjanka K. Schmidt,[14,15] Annegien Broeks,[15] Senno Verhoef,[16] Sten Cornelissen,[15] Kenneth Muir,[17] Artitaya Lophatananon,[17] Sarah Stewart-Brown,[17] Pornthep Siriwanarangsan,[18] Peter A. Fasching,[19,20] Christian R. Loehberg,[20] Arif B. Ekici,[21] Matthias W. Beckmann,[20] Julian Peto,[22] Isabel dos Santos Silva,[22] Nichola Johnson,[23] Zoe Aitken,[22] Elinor J. Sawyer,[24] Ian Tomlinson,[25] Michael J. Kerin,[26] Nicola Miller,[26] Frederik Marme,[27,28] Andreas Schneeweiss,[27,28] Christof Sohn,[27] Barbara Burwinkel,[27,29] Pascal Guénel,[30,31] Thérèse Truong,[30,31] Pierre Laurent-Puig,[32] Florence Menegaux,[30,31] Stig E. Bojesen,[33,34] Børge G. Nordestgaard,[33,34] Sune F. Nielsen,[33,34] Henrik Flyger,[35] Roger L. Milne,[36] M. Pilar Zamora,[37] Jose Ignacio Arias Perez,[38] Javier Benitez,[10,39] Hoda Anton-Culver,[40] Hermann Brenner,[41] Heiko Müller,[41] Volker Arndt,[41] Christa Stegmaier,[42] Alfons Meindl,[43] Peter Lichtner,[44] Rita K. Schmutzler,[45] Christoph Engel,[46] Hiltrud Brauch,[47,48] Ute Hamann,[49] Christina Justenhoven,[47,48] The GENICA Network,[47,48,49,50,51,52,53]

[1]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland 4072, Australia; [2]Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge CB1 8RN, UK; [3]Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK; [4]Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge CB1 8RN, UK; [5]Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki 00029, Finland; [6]Department of Genetics, Queensland Institute of Medical Research, Brisbane, Queensland 4029, Australia; [7]Harvard Medical School and Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA; [8]Department of Epidemiology, Harvard School of Public Health, Boston, MA 02215, USA; [9]Department of Biostatistics, Harvard School of Public Health, Boston, MA 02215, USA; [10]Human Genotyping-CEGEN Unit, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid 28029, Spain; [11]Centre d'innovation Génome Québec et Université McGill, Montréal, QC H3A 0G1, Canada; [12]Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, The University of Melbourne, Melbourne, Victoria 3010, Australia; [13]Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Victoria 3010, Australia; [14]Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, 1066 CX Amsterdam, the Netherlands; [15]Division of Molecular Pathology, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, 1066 CX Amsterdam, the Netherlands; [16]Family Cancer Clinic, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, 1066 CX Amsterdam, the Netherlands; [17]Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK; [18]Ministry of Public Health, Bangkok 10400, Thailand; [19]Department of Medicine, Division of Hematology and Oncology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; [20]University Breast Center Franconia, Department of Gynecology and Obstetrics, University Hospital Erlangen, 91054 Erlangen, Germany; [21]Institute of Human Genetics, Friedrich Alexander University Erlangen-Nuremberg, 91054 Erlangen, Germany; [22]Non-communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK; [23]Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London SW3 6JB, UK; [24]Division of Cancer Studies, NIHR Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust in partnership with King's College London, London SE1 9RT, UK; [25]Welcome Trust Centre for Human Genetics and Oxford Biomedical Research Centre, University of Oxford, Oxford OX3 7BN, UK; [26]Surgery, Clinical Science Institute, Galway University Hospital and National University of Ireland, Galway, Ireland; [27]Department of Obstetrics and Gynecology, University of Heidelberg, 69115 Heidelberg, Germany; [28]National Center for Tumor Diseases, University of Heidelberg, 69120 Heidelberg, Germany; [29]Molecular Epidemiology Group, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [30]INSERM (National Institute of Health and Medical Research), CESP (Center for Research in Epidemiology and Population Health), U1018, Environmental Epidemiology of Cancer Team, 94807 Villejuif, France; [31]University Paris-Sud, UMRS 1018, 94807 Villejuif, France; [32]Université Paris Sorbonne Cité, UMR-S775 INSERM, 75270 Paris Cedex 06, France; [33]Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, Copenhagen, 2730 Herlev, Denmark; [34]Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, University of Copenhagen, Copenhagen, 2730 Herlev, Denmark; [35]Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Copenhagen, 2730 Herlev, Denmark; [36]Genetic & Molecular Epidemiology Group, Human Cancer Genetics Program, Spanish National Cancer Research Centre (CNIO), Madrid 28029, Spain; [37]Servicio de Oncología Médica, Hospital Universitario La Paz, Madrid 28046, Spain; [38]Servicio de Cirugía General y Especialidades, Hospital Monte Naranco, Oviedo 33012, Spain; [39]Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid 28029, Spain; [40]Department of Epidemiology, University of California, Irvine, Irvine, CA 92697, USA; [41]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [42]Saarland Cancer Registry, 66024 Saarbrücken, Germany; [43]Division of Gynaecology and Obstetrics, Technische Universität München, 81675 Munich, Germany; [44]Institute of Human Genetics, Helmholtz Zentrum München - German Research Center for Environmental Health, 85764 Neuherberg, Germany; [45]Division of Molecular Gyneco-Oncology, Department of Gynaecology and Obstetrics, University Cologne, 50931 Cologne, Germany; [46]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, 04107 Leipzig, Germany; [47]Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, 70376 Stuttgart, Germany; [48]University of Tübingen, 72074 Tübingen, Germany; [49]Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [50]Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, 53113 Bonn, Germany; [51]Institute and Outpatient Clinic of Occupational Medicine, Saarland University Medical Center and Saarland University Faculty of Medicine, 66421 Homburg, Germany; [52]Institute for Prevention and Occupational Medicine of the German Social Accident Insurance (IPA), 44789 Bochum, Germany; [53]Institute of Pathology, Medical Faculty of the University of Bonn, 53123 Bonn, Germany;

Kirsimari Aaltonen,[5,54] Päivi Heikkilä,[55] Kristiina Aittomäki,[54] Carl Blomqvist,[56] Keitaro Matsuo,[57] Hidemi Ito,[57] Hiroji Iwata,[58] Aiko Sueta,[57] Natalia V. Bogdanova,[59,60] Natalia N. Antonenkova,[61] Thilo Dörk,[59] Annika Lindblom,[62] Sara Margolin,[63] Arto Mannermaa,[64,65] Vesa Kataja,[65,66] Veli-Matti Kosma,[64,65] Jaana M. Hartikainen,[64,65] kConFab Investigators,[67] Anna H. Wu,[68] Chiu-chen Tseng,[68] David Van Den Berg,[68] Daniel O. Stram,[68] Diether Lambrechts,[69,70] Stephanie Peeters,[71] Ann Smeets,[71] Giuseppe Floris,[71] Jenny Chang-Claude,[72] Anja Rudolph,[72] Stefan Nickels,[72] Dieter Flesch-Janys,[72,73] Paolo Radice,[74,75] Paolo Peterlongo,[74,75] Bernardo Bonanni,[76] Domenico Sardella,[75] Fergus J. Couch,[77] Xianshu Wang,[77] Vernon S. Pankratz,[78] Adam Lee,[79] Graham G. Giles,[12,80] Gianluca Severi,[12,80] Laura Baglietto,[12,80] Christopher A. Haiman,[68] Brian E. Henderson,[68] Fredrick Schumacher,[68] Loic Le Marchand,[81] Jacques Simard,[82] Mark S. Goldberg,[83,84] France Labrèche,[85] Martine Dumont,[82] Soo Hwang Teo,[86,87] Cheng Har Yip,[87] Char-Hong Ng,[87] Eranga Nishanthie Vithana,[88] Vessela Kristensen,[89,90] Wei Zheng,[91] Sandra Deming-Halverson,[91] Martha Shrubsole,[91] Jirong Long,[91] Robert Winqvist,[92] Katri Pylkäs,[92] Arja Jukkola-Vuorinen,[93] Mervi Grip,[94] Irene L. Andrulis,[95,96] Julia A. Knight,[97,98] Gord Glendon,[96] Anna Marie Mulligan,[99,100] Peter Devilee,[101] Caroline Seynaeve,[102,103] Montserrat García-Closas,[23,104,105] Jonine Figueroa,[106] Stephen J. Chanock,[106] Jolanta Lissowska,[107] Kamila Czene,[108] Daniel Klevebring,[108] Nils Schoof,[108] Maartje J. Hooning,[103] John W.M. Martens,[103] J. Margriet Collée,[109] Madeleine Tilanus-Linthorst,[109] Per Hall,[108] Jingmei Li,[110] Jianjun Liu,[110] Keith Humphreys,[108] Xiao-Ou Shu,[91] Wei Lu,[111] Yu-Tang Gao,[112] Hui Cai,[91] Angela Cox,[113] Sabapathy P. Balasubramanian,[113] William Blot,[91,114] Lisa B. Signorello,[91,114] Qiuyin Cai,[91] Paul D.P. Pharoah,[2,4] Catherine S. Healey,[2] Mitul Shah,[2] Karen A. Pooley,[4] Daehee Kang,[115] Keun-Young Yoo,[115] Dong-Young Noh,[115] Mikael Hartman,[116,117] Hui Miao,[117] Jen-Hwei Sng,[116]

[54]Department of Clinical Genetics, University of Helsinki and Helsinki University Central Hospital, Helsinki, 00029, Finland; [55]Department of Pathology, University of Helsinki and Helsinki University Central Hospital, Helsinki, 00029, Finland; [56]Department of Oncology, University of Helsinki and Helsinki University Central Hospital, Helsinki, 00029, Finland; [57]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya 464-8681, Japan; [58]Department of Breast Oncology, Aichi Cancer Center Hospital, Nagoya 464-8681, Japan; [59]Department of Obstetrics and Gynaecology, Hannover Medical School, 30625 Hannover, Germany; [60]Department of Radiation Oncology, Hannover Medical School, 30625 Hannover, Germany; [61]N.N. Alexandrov Research Institute of Oncology and Medical Radiology, 223040 Minsk, Belarus; [62]Department of Molecular Medicine and Surgery, Karolinska Institutet, 171 77 Stockholm, Sweden; [63]Department of Oncology-Pathology, Karolinska Institutet, 171 77 Stockholm, Sweden; [64]Imaging Center, Department of Clinical Pathology, Kuopio University Hospital, 70211 Kuopio, Finland; [65]School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, Biocenter Kuopio, Cancer Center of Eastern Finland, University of Eastern Finland, 70211 Kuopio, Finland; [66]Cancer Center, Kuopio University Hospital, 70211 Kuopio, Finland; [67]Peter MacCallum Cancer Center, Melbourne, Victoria 3002, Australia; [68]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA; [69]Laboratory for Translational Genetics, Department of Oncology, University of Leuven, 3000 Leuven, Belgium; [70]Vesalius Research Center (VRC), VIB, 3000 Leuven, Belgium; [71]Multidisciplinary Breast Center, University Hospital Leuven and KU Leuven, 3000 Leuven, Belgium; [72]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [73]Department of Cancer Epidemiology/Clinical Cancer Registry and Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, 20246 Hamburg, Germany; [74]Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale Tumori (INT), 20133 Milan, Italy; [75]IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, 20139 Milan, Italy; [76]Division of Cancer Prevention and Genetics, Istituto Europeo di Oncologia, 20141 Milan, Italy; [77]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA; [78]Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; [79]Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN 55905, USA; [80]Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Victoria 3053, Australia; [81]Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, HI 96813, USA; [82]Cancer Genomics Laboratory, Centre Hospitalier Universitaire de Québec and Laval University, Québec City, QC G1V 4G2, Canada; [83]Department of Medicine, McGill University, Montreal, QC H3A 1A1, Canada; [84]Division of Clinical Epidemiology, McGill University Health Centre, Royal Victoria Hospital, Montreal, QC H3A 1A1, Canada; [85]Département de médecine sociale et préventive, Département de santé environnementale et santé au travail, Université de Montréal, Montreal, QC H3A 3C2, Canada; [86]Cancer Research Initiatives Foundation, Sime Darby Medical Centre, Subang Jaya, 47500 Selangor, Malaysia; [87]Breast Cancer Research Unit, University Malaya Cancer Research Institute, Faculty of Medicine, University Malaya, 50603 Kuala Lumpur, Malaysia; [88]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore 168751, Singapore; [89]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Radiumhospitalet, 0310 Oslo, Norway; [90]Faculty of Medicine (Faculty Division Ahus), University of Oslo, 0318 Oslo, Norway; [91]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN 37203, USA; [92]Laboratory of Cancer Genetics and Tumor Biology, Department of Clinical Genetics and Biocenter Oulu, University of Oulu, Oulu University Hospital, 90014 Oulu, Finland; [93]Department of Oncology, Oulu University Hospital, University of Oulu, 90014 Oulu, Finland; [94]Department of Surgery, Oulu University Hospital, University of Oulu, 90014 Oulu, Finland; [95]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada; [96]Ontario Cancer Genetics Network, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada; [97]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5T 3M7, Canada; [98]Prosserman Centre for Health Research, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON M5T 3L9, Canada; [99]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON M5S 1A8, Canada; [100]Laboratory Medicine Program, University Health Network, 200 Elizabeth Street, Toronto, ON M5G 2C4, Canada; [101]Department of Human Genetics & Department of Pathology, Leiden University Medical Center, 2300 RC Leiden, the Netherlands; [102]Family Cancer Clinic, Department of Medical Oncology, Erasmus MC-Daniel den Hoed Cancer Center, 3075 EA Rotterdam, the Netherlands; [103]Department of Medical Oncology, Erasmus University Medical Center, 3075 EA Rotterdam, the Netherlands; [104]Division of Genetics and Epidemiology, Institute of Cancer Research and Breakthrough Breast Cancer Research Centre, London SM2 5NG, UK; [105]Division of Breast Cancer Research, The Institute of Cancer Research, London SW3 6JB, UK; [106]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20892, USA; [107]Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center & Institute of Oncology, 02-781 Warsaw, Poland; [108]Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17 177, Sweden; [109]Department of Clinical Genetics, Erasmus University Medical Center, 3008 AE Rotterdam, the Netherlands; [110]Human Genetics Division, Genome Institute of Singapore, Singapore 138672, Singapore; [111]Shanghai Center for Disease Control and Prevention, Shanghai 200336, China; [112]Department of Epidemiology, Shanghai Cancer Institute, Shanghai 200032, China; [113]CRUK/YCR Sheffield Cancer Research Centre, Department of Oncology, University of Sheffield, Sheffield S10 2RX, UK; [114]International Epidemiology Institute, Rockville, MD 20850, USA; [115]Seoul National University College of Medicine, Seoul 110-799, Korea; [116]Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore; [117]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597, Singapore;

Xueling Sim,[118] Anna Jakubowska,[119] Jan Lubinski,[119] Katarzyna Jaworska-Bieniek,[119,120] Katarzyna Durda,[119] Suleeporn Sangrajrang,[121] Valerie Gaborieau,[122] James McKay,[122] Amanda E. Toland,[123] Christine B. Ambrosone,[124] Drakoulis Yannoukakos,[125] Andrew K. Godwin,[126] Chen-Yang Shen,[127,128] Chia-Ni Hsiung,[128] Pei-Ei Wu,[129] Shou-Tung Chen,[130] Anthony Swerdlow,[104,105] Alan Ashworth,[23] Nick Orr,[23,105] Minouk J. Schoemaker,[104] Bruce A.J. Ponder,[3] Heli Nevanlinna,[5] Melissa A. Brown,[1] Georgia Chenevix-Trench,[6] Douglas F. Easton,[2,4] and Alison M. Dunning[2,*]

Analysis of 4,405 variants in 89,050 European subjects from 41 case-control studies identified three independent association signals for estrogen-receptor-positive tumors at 11q13. The strongest signal maps to a transcriptional enhancer element in which the G allele of the best candidate causative variant rs554219 increases risk of breast cancer, reduces both binding of ELK4 transcription factor and luciferase activity in reporter assays, and may be associated with low cyclin D1 protein levels in tumors. Another candidate variant, rs78540526, lies in the same enhancer element. Risk association signal 2, rs75915166, creates a GATA3 binding site within a silencer element. Chromatin conformation studies demonstrate that these enhancer and silencer elements interact with each other and with their likely target gene, *CCND1*.

Introduction

One of the strongest breast cancer associations identified to date via genome-wide association studies (GWASs) is with SNP rs614367 at the 11q13 locus (OR = 1.21; 95% CI 1.17–1.24; $p = 10^{-39}$). This association is restricted to estrogen-receptor-positive (ER^+) tumors.[1,2] SNP rs614367 maps to a 350 kb intergenic region, with *MYEOV1*[3] (MIM 605625) being the nearest centromeric gene and *CCND1* (MIM 168461), *ORAOV1* (MIM 607224), and several genes of the fibroblast growth factor family (*FGF3* [MIM 164950], *FGF4* [MIM 164980], and *FGF19* [MIM 603891]) all lying telomeric, any of which are plausible candidate breast-cancer-susceptibility genes. Although this SNP lies in a gene desert, chromatin modifications suggest that this region contains multiple regulatory elements. Of note, this interval also contains risk SNPs for renal (MIM 144700)[3] and prostate (MIM 176807) cancer.[4–7] Here we report the fine-scale mapping of this locus via 731 SNPs directly genotyped on the custom-designed iCOGS (international Collaborative Oncology Gene-environment Study) Illumina chip together with multiple analyses aimed at exploring the functions of the top independent signals of association with breast cancer.

Material and Methods

Genetic Mapping

Tagging Strategy for Fine-Scale Mapping

In March 2010, when the iCOGS chip was designed, the 1000 Genomes Project (2012) had cataloged 10,358 variants at the 11q13 locus (positions 68,935,424–69,666,272; NCBI build 37 assembly), of which 2,259 had a minor allele frequency (MAF) >0.02. From these, we selected all SNPs having $r^2 > 0.10$ with the originally detected SNP, rs614367, plus a set of SNPs designed to tag all uncorrelated SNPs with $r^2 > 0.8$. After completion of iCOGs genotyping, this initial set was supplemented with a further four SNPs, selected from the October 2010 (Build 37) release of the 1000 Genomes Project, to improve coverage. These were genotyped in two large BCAC (CCHS and SEARCH) studies comprising 12,273 cases and controls, using a Fluidigm array according to manufacturer's instructions. Using the above data, results for all the additional known common variants on the January 2012 release of the 1000 Genomes Project were imputed with IMPUTE version 2.0. Genotypes at 3,674 SNPs were reliably imputed (imputation r^2 score > 0.3) and were analyzed together with the 731 genotyped SNPs—giving a total of 4,405 SNPs within the ~730 kb LD region.

iCOGS Genotyping

Samples were drawn from 50 studies participating in the BCAC: 41 from populations of predominantly European ancestry and 9 of Asian ancestry (unpublished data). Studies were required to provide ~2% of samples in duplicate. All BCAC studies had local human ethical approvals.[8]

Statistical Analysis

For each SNP, we estimated a per-allele log-odds ratio (OR) and standard error by logistic regression, including study and principal components as covariates. Genotype data for all subjects of European ancestry in the study were imputed with the IMPUTE V2.0 software with one phased (January 2012 version of 1000 Genomes project data) and one unphased (CCHS and SEARCH data that were genotyped on the additional four SNPs) reference panel. Association analyses were based on imputed SNPs with estimated MAF > 0 and imputation accuracy $r^2 > 0.3$.

[118]Centre for Molecular Epidemiology, National University of Singapore, Singapore 117597, Singapore; [119]Department of Genetics and Pathology, Pomeranian Medical University, u. Polabska 4, 70-115 Szczecin, Poland; [120]Postgraduate School of Molecular Medicine, Warsaw Medical University, ul. Żwirki i Wigury 61, 02-091 Warsaw, Poland; [121]National Cancer Institute, Bangkok 10400, Thailand; [122]International Agency for Research on Cancer, 69372 Lyon Cedex 08, France; [123]Department of Molecular Virology, Immunology and Medical Genetics, Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA; [124]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY 14263, USA; [125]Molecular Diagnostics Laboratory, IRRP, National Centre for Scientific Research "Demokritos," Athens 15310, Greece; [126]Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, Kansas City, KS 66160, USA; [127]Colleague of Public Health, China Medical University, Taichong 40402, Taiwan, ROC; [128]Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan, ROC; [129]Taiwan Biobank, Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan, ROC; [130]Department of Surgery, Changhua Christian Hospital, Changhua City, Changhua county 500, Taiwan, ROC

[131]These authors contributed equally to this work and are listed in random order

*Correspondence: amd24@medschl.cam.ac.uk

http://dx.doi.org/10.1016/j.ajhg.2013.01.002.

Conditional analyses were performed to identify SNPs independently associated with the phenotype in question. To identify the most parsimonious model, all SNPs with a p value <0.0001 and MAF >0.02 in the single SNP analysis were included in forward selection regression analyses with penalty k = 10 in the step function in R. Haplotype-specific ORs were estimated by in-house methods based on the tagSNPs program[9] and haplostats.[10] Study and principal components were included as covariates. The contribution of 11q13 variants to the familial risk of breast cancer was estimated with the formula $\log(\lambda_L)/\log(\lambda_0)$. Here λ_L is the familial relative risk to daughters of individuals with breast cancer explained by the locus under an additive model, given by

$$\lambda_L = \frac{\sum_{k=1}^{K} p_k \psi_k^2}{\left(\sum_{k=1}^{K} p_k \psi_k\right)^2},$$

where $_K$ is the number of alleles or haplotypes, p_K is the frequency of the kth allele (haplotype), and ψ_K is the per-allele (per-haplotype) relative risk. λ_0 is the overall familial relative risk to degree relatives of individuals with breast cancer, assumed to be 2. For ER-positive breast cancer, the same overall familial relative risk ($\lambda_0 = 2$) was assumed. p values for evaluation of differences in cyclin D1 protein levels by SNP genotype were calculated with χ^2 test or Fisher's exact test by SPSS v18.0.2 (SPSS, Inc.).

CCND1 Protein and Gene Expression

Tissue microarrays (TMAs) were previously constructed on 1,348 invasive breast tumors from the HEBCS study and processed as described,[11] including four cores (diameter 0.6 mm) of the most representative area from each formalin-fixed and paraffin-embedded breast cancer specimen. For cyclin D1 protein levels, TMA slides were stained with cyclin D1 (Novocastra) antibodies (diluted 1:20). Cyclin D1-positive cells were counted in one high-power field (objective 40×) in each of the four cores on TMA. Only unequivocal positive nuclear staining was accepted as a positive reaction. A minimum of 200 breast cancer cells was counted in each tumor. The result was the percentage of all positive cells from the entire number of breast cancer cells counted from the four biopsies. Tumors with expression levels below 1% were considered as negative and above 1% as positive (Figure S1). In total, 644 individuals with breast cancer had both TMA information and genotypic data (genotyped in iCOGS), and 512 of the breast tumors were ER positive. The correlation of SNP rs554219 and cyclin D1 protein levels was examined in the presence of SNP rs75915166 common homozygotes.

To perform eQTL analyses with data from breast cancer cases in The Cancer Genome Atlas Project, we downloaded data from 382 TCGA breast cancer cases from the TCGA data portal. We obtained germline SNP data by using the birdseed genotype calls from the Level 2 Affymetrix 6.0 arrays on peripheral blood. We obtained the normalized log2 tumor gene expression profiling data from the level 2 Agilent microarray data. Germline SNPs were imputed to the 1000 Genomes data set by PLINK. We assessed the association of the germline risk alleles with tumor *CCND1* expression in the 301 ER^+ breast cancers.

Cell Lines

Breast cancer cell lines MDAMB415, CAL51, MCF7, MDAMB231, PMC42, and HCC1954 were grown in DMEM medium with 10% FCS and antibiotics under standard conditions, while T47D was grown in RPMI medium with the same supplements. Where relevant, cell lines were genotyped with fluorescent 5′exonulease assay (TaqMan) and the ABI prism 7900 Sequence Detection System (PE Biosystems) in a 384-well format.

Chromatin Interaction Analysis by Paired-End Tag Sequencing Analysis

PETCluster data files representing two ChIA-PET libraries (IHM001F and IH015F) prepared with ER antibodies in MCF7 cells were downloaded from the ChIA-PET Browser. Analysis of interactions within the putative regulatory element 1 (PRE1; NCBI build 36 chr11: 69,036,648–69,042,291) was conducted in R/Bioconductor. Closest neighbor genes were identified with the Bioconductor package ChiPpeakAnno.

Chromatin Conformation Capture

Chromatin conformation capture (3C) libraries were generated with HindIII and DpnII as described previously.[12] 3C interactions were quantitated by real-time PCR with primers designed within the restriction fragments of interest (Table S4). qPCR was performed on a RotorGene 6000 using MyTaq HS DNA polymerase (Bioline) with the addition of 5 mM of Syto9, annealing temperature of 66°C, and extension of 30 s. HindIII 3C analysis of MCF7 cells was performed in two independent experiments. DpnII 3C analysis of MCF7 cells was performed in three independent experiments, and T47D, MDAMB231, and CAL51 analysis was performed in two independent experiments. Each experiment was quantified in triplicate. Two BAC clones (RP11-156B3 and RP11-378K8) covering the *11q13/CCND1* region were used to create an artificial library of ligation products in order to normalize for PCR efficiency. Data were normalized to the signal from the BAC clone library and, between cell lines, by reference to a region within *GAPDH*. All qPCR products were electrophoresed on 2% agarose gels, gel purified, and sequenced to verify the 3C product.

Plasmid Construction and Luciferase Assays

A *CCND1* promoter-driven luciferase reporter construct was generated by inserting a 2,746 bp fragment containing the *CCND1* promoter into the KpnI and HindIII sites of pGL3-basic. To assist cloning, AgeI, NheI, and SbfI sites were inserted into the BamHI and SalI sites downstream of luciferase. A 3,340 bp fragment containing PRE1 was inserted into the AgeI and SbfI sites or a 955 bp fragment containing the putative regulatory element 2 (PRE2) was inserted into the BamHI and SalI sites downstream of luciferase. Individual SNPs were incorporated into PRE1 and PRE2 via overlap extension PCR. PRE2 was incorporated into PRE1 containing constructs by inserting PRE2 into NheI and SalI sites downstream of PRE1. All constructs were sequenced to confirm variant incorporation (AGRF, Australia). Primers used to generate all constructs are listed in Table S4.

MCF7, T47D, or CAL51 cells were transfected with equimolar amounts of luciferase reporter plasmids and 50 ng of pRLTK with Lipofectamine 2000. The total amount of transfected DNA was kept constant per experiment by adding carrier plasmid (pUC19). Luciferase activity was measured 24 hr post-transfection by the Dual-Glo Luciferase Assay System on a Beckman-Coulter DTX-880 plate reader. To correct for any differences in transfection efficiency or cell lysate preparation, *Firefly* luciferase activity was normalized to *Renilla* luciferase.

The activity of each test construct was calculated relative to *CCND1* promoter construct, the activity of which was arbitrarily defined as 1. For knockdown experiments, MCF7 cells were cotransfected with the relevant luciferase reporter plasmids and either 100 nM of Dharmacon SMARTpool siRNA or shRNA plasmids with Lipofectamine 2000 (Invitrogen). 48 hr after transfection, the luciferase activity was performed as described above.

siRNA Knockdown

ON-TARGET*plus* SMARTpool siRNAs for *GABPA* (MIM 600609; L-011662-00) and *GATA3* (MIM 131320; L-003781-00) and nontargeting siRNA (D-001810-10-20) were purchased from Dharmacon (Thermo Scientific). Two *ELK4* (MIM 600246) small hairpin shRNA constructs corresponding to two independent shRNA sequences have been described previously.[13]

Quantitative PCR

Total RNA was extracted with Trizol (Invitrogen) and reverse transcribed with random hexamers and SuperScriptIII (Invitrogen) according to manufacturers' instructions. qPCR was performed on a RotorGene 6000 (Corbett Research) with TaqMan Gene Expression assays (Hs00360812_m1 for ELK4, Hs00231122_m1 for GATA3, and Hs01022023_m1 for GABPA) and TaqMan Universal PCR master mix. All reactions were normalized against *β-glucuronidase* (MIM 611499; Cat# 4326320E).

Electrophoretic Mobility Shift Assay

Small-scale nuclear extracts and bandshifts were carried out as previously described[14] and oligonucleotide sequences used in the assays are listed in Table S4. Antisera were obtained from Santa Cruz Biotech: rabbit polyclonal antisera were used against USF1 (sc229x), USF2 (sc862x), SP1 (sc14027x), GABPA (sc228x), and ELK4 (sc13030x). GATA3 (sc268x) was detected with mouse monoclonal antibodies. Competitor oligonucleotides were used at 10-, 30-, and 100-fold molar excess.

Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) experiments were carried out as previously described.[15] In brief, cells were cross-linked in 1% formaldehyde for 10 min at room temperature before harvesting and washing in PBS, 1× protease inhibitors (PI; Roche). Cells were lysed and washed to remove the cytoplasm and the nuclei resuspended in LB3 (10 mM Tris-HCl [pH 8], 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium-deoxycholate, 0.5 N-laurylsarcosine, 1× PI) and sonicated for 15 cycles (30 s on, 30 s off, 4°C, high setting) on a Diagenode Bioruptor. Lysates were cleared at 14,000 × *g* for 10 min and incubated with 10 μg of antibody and 10 μl of Protein-G beads (Dynal) for 12 to 18 hr at 4°C (antibodies as for EMSA). Beads were washed in RIPA buffer and DNA isolated by standard methods (-QIAGEN). DNA was quantitated with Quant-IT and equal amounts of precipitate and input used in RT-PCR reaction with primers given in Table S4. Allele-specific PCR was carried out with TaqMan Genotyping Assays (predesigned assays, ABI). All values obtained are normalized to input and enrichment is given relative to the negative *CCND1* control.[16] Each ChIP has yielded similar results in at least two independent experiments. The error bars denote the standard deviation in three technical repeats.

In Silico Analysis of GATA3 Binding

To examine the potential of GATA3 to bind rs75915166, we calculated the enrichment of the 6 bp motif overlapping the core of the position weight matrix (PWM) of the GATA3 motif around this SNP in the 75 bp either side of all GATA3 ChIP-seq peaks,[17] compared to random genomic sequences of the same length. Furthermore, to examine the differential enrichment underneath GATA peaks for the motif containing the A allele over the T allele, the relative enrichment between proportions of peaks containing these motifs was compared to relative enrichments observed within 100,000 sets of random intergenic sequences. Enrichment of the motif containing the A allele compared to T under GATA peaks was compared to the bootstrapped distribution of relative enrichments within random genomic regions yielding a p value of 0.0147.

Results

Case-Control Studies

The original GWAS-associated SNP, rs614367, tags a linkage disequilibrium (LD) block of 683 kb spanning chromosome 11 positions 68,935,424–69,666,272 (NCBI build 37 assembly), as defined by the furthest telomeric and centromeric SNPs displaying detectable correlation ($r^2 > 0.10$) with rs614367. With data from the 1000 Genomes Project, we cataloged 10,358 variants in the region and selected a subset of these to cover the entire region (see Material and Methods). Of these, 731 SNPs were successfully designed and genotyped on the iCOGS chip in 41 case-control studies from populations of European ancestry (89,050 subjects) and 12,893 subjects from 9 case-control studies of Asian ancestry within BCAC (Supplemental Data). Genotypes of all other known variants in the locus were imputed in the European studies by using known genotypes in combination with data from the 1000 Genomes Project. 3,674 SNPs were reliably imputed (imputation r^2 score > 0.3) and were considered for further analysis together with the 731 genotyped SNPs.

Based on data from all European studies, 204 genotyped or imputed SNPs were convincingly associated with overall risk of breast cancer (p values 10^{-5} to 10^{-64}, Table S1). Stratification by tumor ER status confirmed that all associations were with ER^+ disease with no significant evidence for any SNPs associated with ER^- tumors (Table S1). Thus, all further analyses were confined to risks of ER^+ disease. A Manhattan plot of all considered SNPs at this locus (Figure 1A) demonstrates a complex pattern of association.

Stepwise Logistic Regression Reveals Multiple Independent Signals

To dissect this pattern of associations, all genotyped SNPs displaying evidence for association with ER^+ disease (87 SNPs, $p < 10^{-4}$) were included in forward stepwise regression models. The most parsimonious model included three independent SNPs significant at $p < 10^{-4}$ (Table 1,

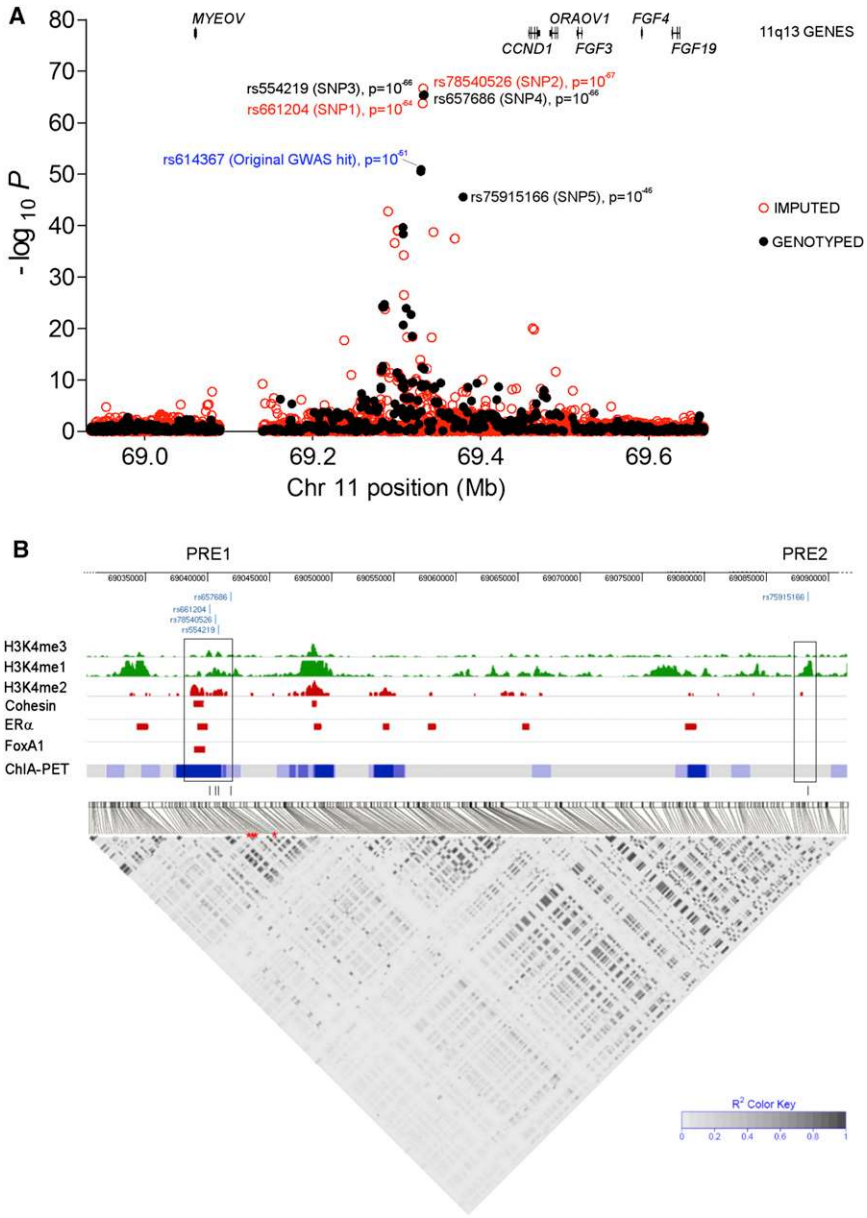

Figure 1. Genetic Mapping and Epigenetic Landscape at the 11q13 Locus

(A) Manhattan plot of the 11q13 susceptibility locus for breast cancer. Genotyped and imputed SNPs are plotted based on their chromosomal position on the x axis and their overall p values (log10 values) from the European BCAC studies on the y axis. The six genes present in the region are indicated in black.

(B) Epigenetic and transcriptional landscape at the 11q13 risk locus for breast cancer in human mammary epithelial cells (HMECs). Green and red histograms denote ChIP-seq data from HMECs (ENCODE) and MCF7 cells stimulated with estrogen (GEO #GSM594606); blue denotes a heat map of ERα ChIA-PET data from MCF7s treated with estradiol.[16] Red bars denote cohesin (ArrayExpress; #E-TABM-828), ERα (GEO #GSM365926), and FoxA1[18] ChIP-seq data from MCF7 cells. Abbreviations are as follows: PRE1, putative regulatory element 1 that contains SNPs 1–4; PRE2, putative regulatory element 2 that contains SNP5. Below depicts the pattern of linkage disequilibrium with data from the BCAC population, where white represents $r^2 = 0$ and black $r^2 = 1$. Red stars denote the positions of SNPs 1–4 in the linkage block.

Figure 1A). These were (1) SNP rs554219 (OR per minor allele = 1.33; 95% CI 1.28–1.37; p value 10^{-66}; conditional p value 3.7×10^{-26}); (2) SNP rs75915166 (OR per minor allele = 1.38; 95% CI 1.32–1.44; p value 2.7×10^{-46}; conditional p value 2.7×10^{-8}); and (3) rs494406 (OR per minor allele = 1.07; 95% CI 1.05–1.11; p value 3.7×10^{-9}; conditional p value 2.6×10^{-6}) (Figure 1A).

Variants for subsequent functional analysis were selected, on the basis of the above analysis of genotyped SNPs, according to the following criteria: assuming a single causative variant for each of the independent signals, we calculated the likelihood ratio of each SNP relative to best independent signal with which it was correlated. SNPs with a likelihood of <1:100 compared with the most significant SNP for each signal were excluded from consideration as being potentially causative. For signal 1, four SNPs clustered in a 20 kb region (position: 69,320,000–69,340,000 build 37) remained after this exclusion process: rs661204 (SNP1), rs78540526 (SNP2), rs554219 (SNP3), and rs657686 (SNP4). For signal 2, only SNP rs75915166 (SNP5) remains; all other SNPs correlated with this one had much less significant effects. Of note rs75915166 is partially correlated with the signal 1 SNPs (r^2 with rs554219 = 0.61) but conditional analysis indicates it is clearly independent (p value 5×10^{-8} after conditioning on rs554219). For signal 3, four SNPs (rs494406, rs585568, rs593679, rs679162) remain as potentially causal, but these are associated with much smaller effect sizes (Table 1). Further investigations were thus focused on the five SNPs (SNP1–SNP5; listed above) for which there was strongest evidence for likely causation.

When the stepwise regression was repeated, after imputation of all SNPs in the locus to the January 2012 release of the 1000 Genomes data, the most parsimonious model included (1) SNP rs78540526 (SNP2; conditional p value 4×10^{-10}); (2) SNP rs554219 (SNP3; conditional p value 9×10^{-8}); and (3) a newly discovered variant at chromosome 11: SNP rs12575120 (conditional p value 3×10^{-6}). No overall evidence of heterogeneity was observed for the genotyped SNPs, which we selected for functional analysis, among the European or Asian studies

Table 1. Association of the Three Independent SNPs Is Strictly Confined to ER$^+$ Breast Cancer

Signal	SNPs	Chromosome Position[a]	MAF[b]	OR ER$^+$ 95% CI	p Trend	OR ER$^-$ 95% CI	p Trend	p Value in Logistic Regression
Signal 1	rs554219	69331642	0.123	1.33 (1.28–1.37)	5.64×10^{-66}	1.02 (0.97–1.08)	0.47	3.7×10^{-26}
	rs657686	69332670	0.122	1.33 (1.29–1.37)	4.07×10^{-66}	1.02 (0.97–1.08)	0.44	
Signal 2	rs75915166[c]	69379161	0.059	1.38 (1.32–1.44)	2.70×10^{-46}	1.06 (0.99–1.15)	0.11	2.7×10^{-8}
Signal 3	rs494406	69344241	0.255	1.07 (1.05–1.11)	3.73×10^{-9}	1.02 (0.98–1.07)	0.34	2.6×10^{-6}
	rs585568	69345336	0.255	1.07 (1.05–1.11)	4.96×10^{-9}	1.02 (0.98–1.06)	0.35	
	rs679162	69344477	0.255	1.07 (1.05–1.11)	5.15×10^{-9}	1.02 (0.98–1.06)	0.36	
	rs593679	69342650	0.255	1.07 (1.05–1.11)	3.76×10^{-9}	1.02 (0.98–1.06)	0.36	
Previously reported GWAS variant	rs614367	69328764	0.161	1.26 (1.22–1.30)	1.28×10^{-51}	1.02 (0.97–1.08)	0.40	–

These three SNPs (rs554219, rs75915166, and rs494406) remain in a forward stepwise logistic regression analysis that included all associated SNPs with ER$^+$ ($p < 0.0001$) and MAF > 0.02.
[a]Build 37.
[b]MAF in controls.
[c]Also named pos69088342 in build 36.

($p > 0.08$) (Figure S2). The minor allele frequencies (MAFs) of these SNPs are much rarer in Asian populations than in Europeans (rs554219 and rs657686, MAF = 0.017; rs75915166 and rs661204, MAF < 0.01) although SNP rs78540526 appeared to be monomorphic in Asians (Table S3). Despite this, the SNPs with detectable minor alleles in Asians have risk estimates for ER$^+$ tumors consistent with those in Europeans (rs554219 [SNP3]: OR = 1.64; 95% CI 1.27–2.11; p value = 1.3×10^{-4}; rs657686 [SNP4]: OR = 1.61; 95% CI 1.25–2.07; p value = 2.2×10^{-4}; rs75915166 [SNP5], OR = 1.42, p value = 3.6×10^{-2}). These significant associations, despite the rarity of the minor alleles in Asian populations, provide further support that these SNPs may have directly causative effects.

Three Distinct Haplotypes Confer Increased Risks with Different Magnitudes

We conducted haplotype analysis with the five SNPs, selected above, which define four common haplotypes (Table 2). Haplotype H1, carrying the risk alleles of SNPs 1, 3, and 4, is associated with a significant increase in ER-positive breast cancer risk over the commonest haplotype (H0) (OR = 1.16, 95% CI 1.12–1.21, p value = 1.6×10^{-8}), and a second, rarer haplotype (H2) carrying risk alleles of SNPs 1, 2, 3, and 4 is associated with a significantly greater risk (OR = 1.39, 95% CI 1.30–1.50, p value = 1.25×10^{-20}). A third haplotype, carrying all five risk alleles, is associated with the highest risk estimate (OR = 1.44, 95% CI 1.38–1.51, p value = 1.22×10^{-55}) although this was not significantly higher than that of H2.

The Minor G Allele of SNP rs554219 May Be Associated with Reduced Cyclin D1 Protein Levels

To explore the target gene of the functional SNPs, we looked for associations of SNPs 1–5 with gene expression in human tissue. We observed no evidence for association of any of these SNPs (1) with the local genes (*MYEOV1, CCND1*, *ORAOV1*, *FGF3*, *FGF4*, and *FGF19*) in RNA from 40 normal breast tissue samples or (2) with *CCND1* expression in 300 ER$^+$ tumors from the TCGA project—but our power to detect such associations in these samples was low. In a set of 448 ER$^+$ breast tumors from the Helsinki Breast Cancer Study (HEBCS), signal 1 SNP rs554219 (SNP3) showed borderline evidence for association with differences in cyclin D1 protein levels (determined by immunohistochemistry) in a recessive model (p = 0.037) but no linear trend was visible (p = 0.69, Table 3). Homozygotes for the risk G allele associated with reduced cyclin

Table 2. Haplotype Analysis across the BCAC Studies

Haplotypes	rs661204 (SNP1)	rs78540526 (SNP2)	rs554219 (SNP3)	rs657686 (SNP4)	rs75915166 (SNP5)	Haplotype Frequency	OR	p Value
1	1	0	1	1	0	0.048	1.16 (1.12–1.21)	1.6×10^{-8}
2	1	1	1	1	0	0.025	1.39 (1.30–1.50)	1.25×10^{-20}
3	1	1	1	1	1	0.062	1.44 (1.38–1.51)	1.22×10^{-55}
All others	rare					0.005	0.99 (0.83–1.18)	0.90

Each haplotype was compared to the ancestral haplotype carrying the common alleles of all five SNPs. SNPs rs661204, rs554219, and rs657686 are perfectly correlated with each other and hence always inherited together.

Table 3. Association of SNP3 rs554219 with Cyclin D1 Protein Levels

SNP rs554219 Genotype	Cyclin D1 Protein Levels		
	Total (n = 448)	Negative (n = 52)	Positive (n = 396)
C/C	320 (71.4)	38 (73.1)	282 (71.2)
C/G	121 (27.0)	11 (21.2)	110 (27.8)
G/G	7 (1.6)	3 (5.8)	4 (1.0)

Protein levels (binary: negative versus positive) detected by immunohistochemistry in 448 ER^+ tumors from cases in the HEBCS. All included cases were homozygous for the common allele of SNP rs75915166 (SNP5). p values were calculated as 0.024 by heterogeneity test, 0.69 by chi-square test for trend, and 0.037 by chi-square test for recessive model.

D1 staining. To avoid possible interference of the second independent risk variant, this analysis was carried out only in samples homozygous for the common allele of SNP rs75915166 (SNP5). After adjustment for rs554219, SNP5 showed no significant correlation with cyclin D1 protein levels.

The Strongest Candidate Causal SNPs Map to Two Putative Regulatory Elements that Distally Regulate the *CCND1* Promoter

Regulatory elements such as transcriptional enhancers and silencers can be identified by transcription factor (TF) occupancy and distinct chromatin marks such as mono- and dimethylation of histone 3 lysine 4 (H3K4Me1 and H3K4Me2), which mark active promoters and enhancers.[18,19] We used available ChIP-seq data from MCF7 cells for H3K4Me1, H3K4Me2, and selected TFs to determine whether SNPs 1–5 fall within putative transcriptional regulatory elements. Signal 1 SNPs (1–4) cluster within a 1.7 kb LD block that falls in a putative regulatory element (PRE) (PRE1; Figure 1B) flanked by H3K4Me1 and H3K4Me2 marks, and signal 2 SNP5 (rs75915166) lies in a second PRE (PRE2; Figure 1B). Notably, PRE1 binds ER alpha (ERα), which is consistent with the associations being confined to ER^+ tumors.[2]

According to ChIA-PET (chromatin-interaction analysis with paired-end tag sequencing) data generated by Fullwood et al.,[20] PRE1 is a hotspot for ER-bound chromatin interactions (Figure 1A), suggesting that PRE1 regulates distal genes by participating in long-range chromatin interactions. Consistent with this, PRE1 also binds cohesin, a DNA binding protein shown to be important in tethering long-range chromatin interactions and FOXA1, a pioneer factor that initiates ERα-chromatin binding[21,22] (Figure 1B). Mining of the ERα ChIA-PET data identified several genomic regions participating in long-range chromatin interactions with PRE1 in independent biological replicates (Table S3). The *CCND1* promoter, located approximately 125 kb downstream, was the only gene promoter that reproducibly interacts with PRE1, suggesting that PRE1 may be involved in regulating *CCND1* expression. Notably, the ChIA-PET data shows that PRE1 also interacts frequently with the terminator region of *CCND1*, which contains a previously reported enhancer (enh2) of *CCND1*.[16]

Via chromosome conformation capture (3C), we confirmed that PRE1 frequently interacts with the *CCND1* promoter and terminator in ERα-positive MCF7 and T47D cells (Figure 2A). Furthermore, PRE2 also frequently interacts with the *CCND1* promoter (Figure 2B). With DpnII 3C libraries we mapped the PRE1/*CCND1* promoter interaction to two adjacent DpnII fragments within PRE1 spanning 1.5 kb. Interestingly, this interaction was present in the two ERα-positive cell lines, MCF7 and T47D, but greatly reduced in ERα-negative CAL51 and MDAMB231 cell lines (Figure 2C). Of note, SNP rs661204 (SNP1) lies within the restriction fragment shown to be involved in tethering the interaction. However, allele-specific 3C on MDAMB415 cells, a cell line heterozygous for this SNP, revealed that this SNP had no significant effect on chromatin looping (Figure S3). Mapping of the PRE2/*CCND1* promoter interaction, with the same DpnII 3C libraries, showed that PRE2 frequently interacts with the *CCND1* promoter in MCF7, T47D, and MDAMB231 cells but not in CAL51 cells (Figure 2C), suggesting that this interaction can occur in a cell-specific manner independent of ERα. We also detected long-range chromatin interactions between PRE1 and PRE2 in MCF7 and T47D cells, suggesting that these two regulatory elements may cooperate to regulate *CCND1* expression (Figure 2A).

Three of the Five Candidate SNPs Affect the Regulatory Capability of PRE1 and PRE2 on the *CCND1* Promoter

By using luciferase reporter assays in MCF7 cells, we demonstrated that PRE1 is able to act as a strong transcriptional enhancer, leading to a 40-fold increase in *CCND1* promoter activity (Figure 3A), whereas PRE2 ablated *CCND1* promoter activity (Figure 3C), acting as a silencer. A similar effect was also observed in T47D and CAL51 cells (Figure S4) albeit to a lesser extent (6-fold in T47D and 1.8-fold in CAL51 cells). To examine whether SNPs (1–4) affect the enhancer activity of PRE1, we generated reporter constructs containing the minor risk alleles of these SNPs. Significantly, in MCF7 cells the minor alleles of SNPs 2 and 3 (rs78540526 and rs554219) almost completely abolished PRE1 enhancer activity whereas SNPs 1 and 4 (rs661204 and rs657686) had only a minor or no effect (Figure 3A). In T47D and CAL51 cells, similar activities were observed (Figure S4). Consistent with ERα ChIP-seq data (Figure S5), we find that PRE1 is estrogen inducible. This response is not affected by the different alleles of the four SNPs (Figure 3B). Because the silencer strongly represses transcriptional activity, any additional repressive effect of PRE2 SNP rs75915166 (SNP5) would not be readily observed (Figure 3C). We therefore cloned the PRE1 enhancer into the PRE2 constructs to increase

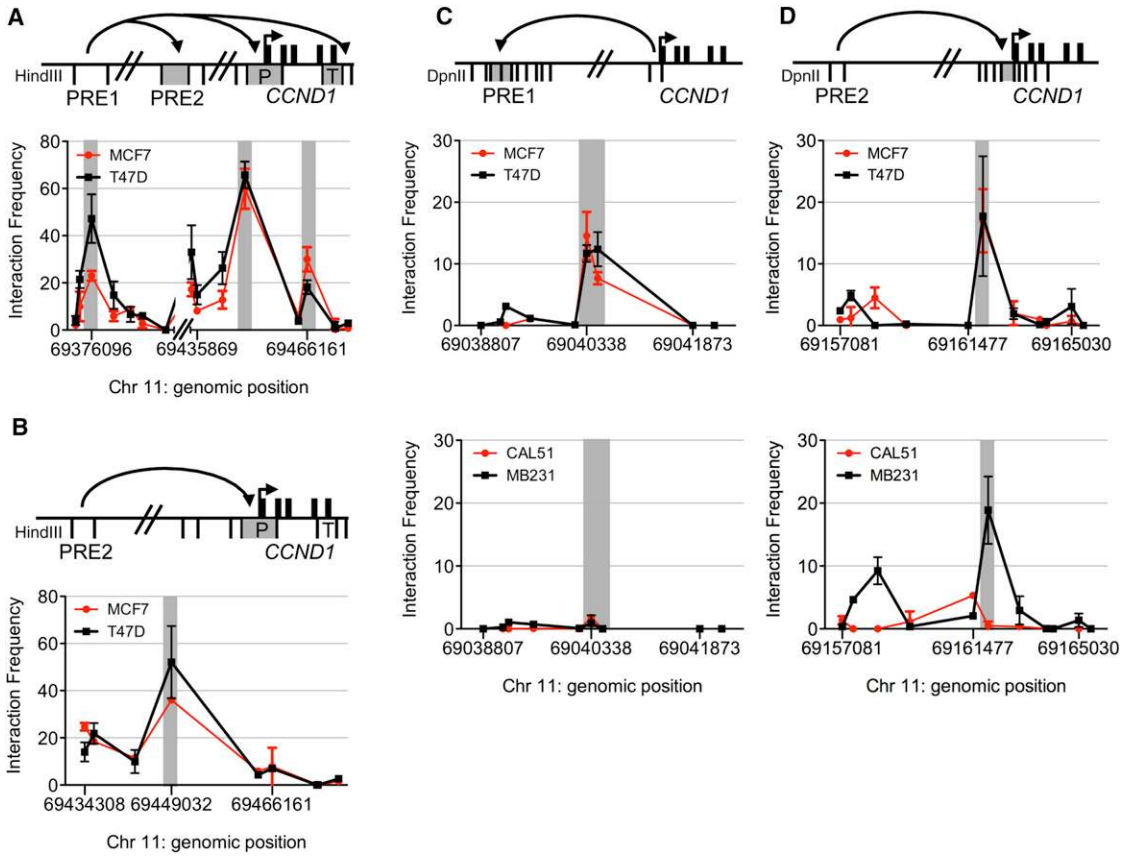

Figure 2. Long-Range Chromatin Interactions of the 11q13 Risk Regions with *CCND1* in Breast Cancer Cell Lines
(A and B) 3C interaction profiles between PRE1 and/or PRE2, the *CCND1* promoter (P), and terminator (T) regions. 3C libraries were generated with HindIII, with the anchor points set at either PRE1 (A) or PRE2 (B). Grey bars depict the position of the target sites and matches them to the cartoons above each panel.
(C) Fine-mapped 3C interaction profiles between the PRE1 and fragments spanning the *CCND1* promoter in ER^+ (MCF7 and T47D) and the ER^- (CAL51 and MDAMB231) cell lines. 3C libraries were generated with DpnII and anchor point is set at the *CCND1* promoter.
(D) Fine-mapped 3C interaction profiles between the *CCND1* promoter and fragments spanning PRE2. Anchor point is set at PRE2. A representative graph of at least two biological replicates is shown.
Error bars represent SD. Physical maps of the regions interrogated by 3C are shown above (not to scale).

luciferase levels (Figure 3D; PRE1+2; PRE1+2-S5). Importantly, we find that in the context of the PRE1 common alleles, the minor allele of SNP5 significantly increased the strength of the silencer (Figure 3D).

ELK4 and GATA3 Mediate the Effects of PRE1 SNP rs554219 and PRE2 SNP rs75915166, Respectively

We used electrophoretic mobility shift assays (EMSA) to examine protein-DNA interaction for SNPs 1–5. All five SNPs displayed TF binding that was allele specific in four cases (Figures 4 and S6A). Competition with known TF binding sites suggested the identity of each of the bound proteins (data not shown), which was confirmed in supershift experiments (Figure 4). Inclusion of antisera in the binding reaction established that the common alleles of SNPs 1 and 2 (rs661204, rs78540526) preferentially bind USF1 and USF2. The common allele of SNP3 (rs554219) is bound by ELK4 and GABPA, whereas the minor allele of SNP5 (rs75915166) interacts specifically with GATA3. A high-mobility complex bound by the oligonucleotide containing SNP5 is independent of allelic status and therefore unlikely to be relevant to cancer risk (Figure S6B). To assess occupancy of the different SNPs in vivo, allele-specific ChIP were carried out by TaqMan assays. Little or no enrichment was detected for USF1 or USF2 on SNPs 1 and 2 and no allelic discrimination was observed (Figure S7). However, in an ELK4 ChIP assay for SNP3 (rs554219), which mediates one of the strongest effects in the transcriptional assays, the common allele (C) shows a 7.7-fold enrichment over a negative control and a 7.1-fold enrichment over the risk allele, indicating that this site is occupied in an allele-specific manner in vivo (Figures 5A and S8). GABPA, which binds the C allele of this site in EMSAs, does not bind this site in vivo as shown by ChIP assays (only 2.5-fold enrichment versus 160-fold enrichment of a positive control; Figure S9). The importance of ELK4 binding was confirmed in cotransfection assays that show that two independent siRNAs against ELK4 reduce enhancer activity of wild-type enhancer, but not of the enhancer containing the rare allele of SNP3

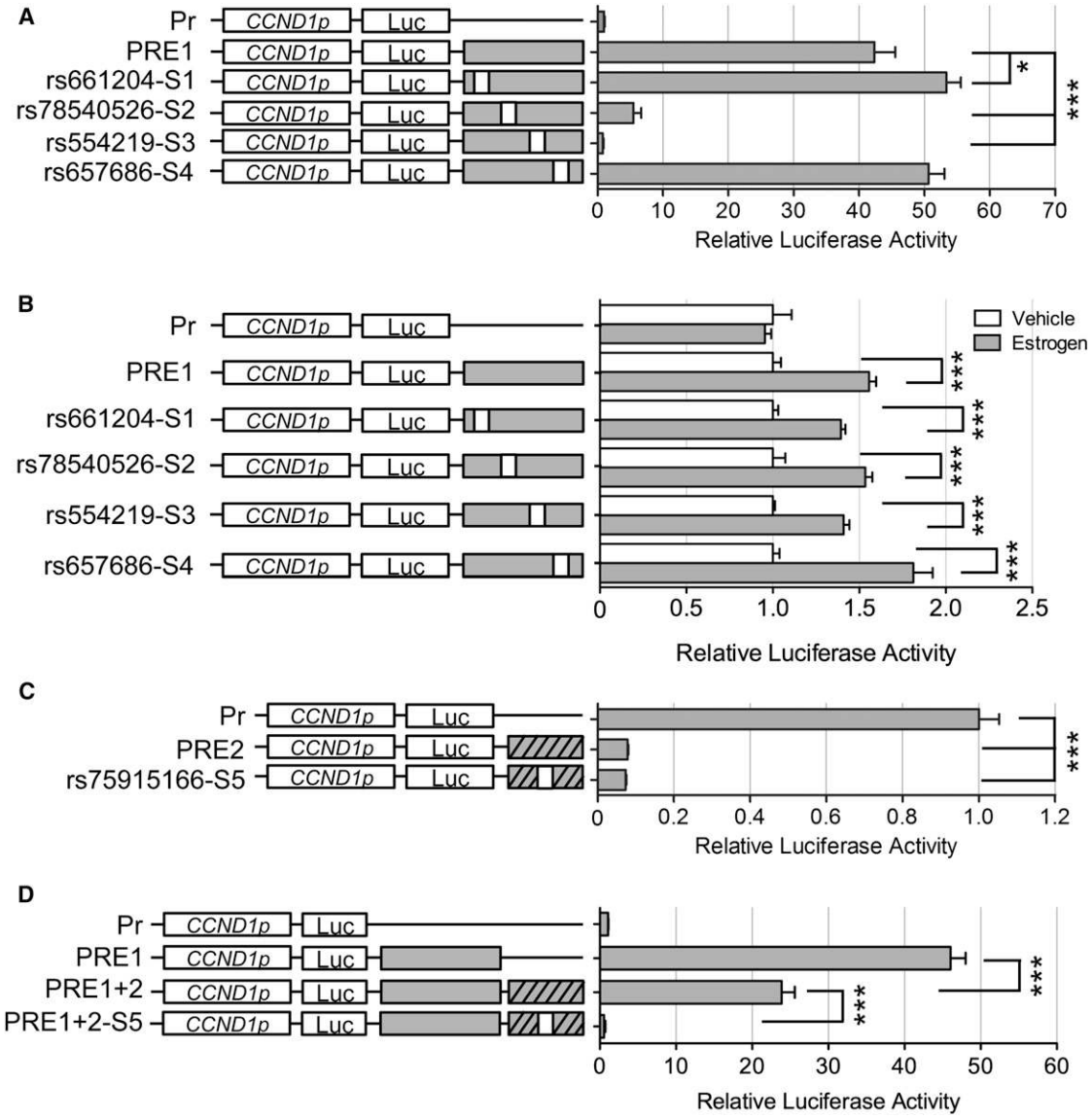

Figure 3. Luciferase Reporter Assays in MCF7 Cells Demonstrating Regulatory Activity at the 11q13 Risk Locus for Breast Cancer
(A and C) PRE1 and PRE2 were cloned downstream of a *CCND1* promoter-driven luciferase reporter (Pr) with and without SNPs 1–5 (rs661204, S1; rs78540526, S2; rs554219, S3; rs657686, S4; rs75915166, S5). MCF7 cells were transiently transfected with each of these constructs and assayed for luciferase activity after 24 hr.
(B) MCF7 cells were transiently transfected with PRE1-luciferase reporter constructs, pretreated with 10 nM ICI 182780 for 24 hr, and then stimulated with estradiol (100 nM) or vehicle for 24 hr. Luciferase activity was normalized to the activity of the vehicle-treated cells.
(D) PRE1 luciferase constructs were generated containing PRE2 with and without SNP5 (PRE1+2; PRE1+2-S5). MCF7 cells were transiently transfected with each of these constructs and assayed for luciferase activity after 24 hr. Representative graphs are shown from at least two independent experiments.
Error bars denote SD from one experiment performed in triplicate. p values were determined with a two-tailed t test. *$p < 0.05$, ***$p < 0.001$.

(rs554219; Figure 5B), further strengthening the conclusion that ELK4 is an important mediator of enhancer function.

GATA3 in vivo binding could not be assessed because our panel of 80 breast cancer cell lines did not include any ER^+ cell lines carrying the minor allele of rs75915166 (SNP5, MAF = 3.5%). However, with a bioinformatics approach, we demonstrated that sequences identical to the sequence surrounding this SNP are enriched under GATA3 ChIP-seq signals ($p < 10^{-6}$) and, importantly, this enrichment is higher for the risk A allele than the common C allele ($p = 0.015$) (Figure 5C), suggesting that our in vitro binding results (Figure 4) are replicated in vivo. Consistent with this, we find that the introduction of the nonbinding C allele into the core consensus motif strongly reduces motif enrichment. We again confirmed the functional importance of GATA3 by using RNAi in luciferase cotransfection assays: a smart pool of siRNA against *GATA3* increases transcription in the presence of the minor allele, which binds GATA3 (Figure 4) but has no effect on the

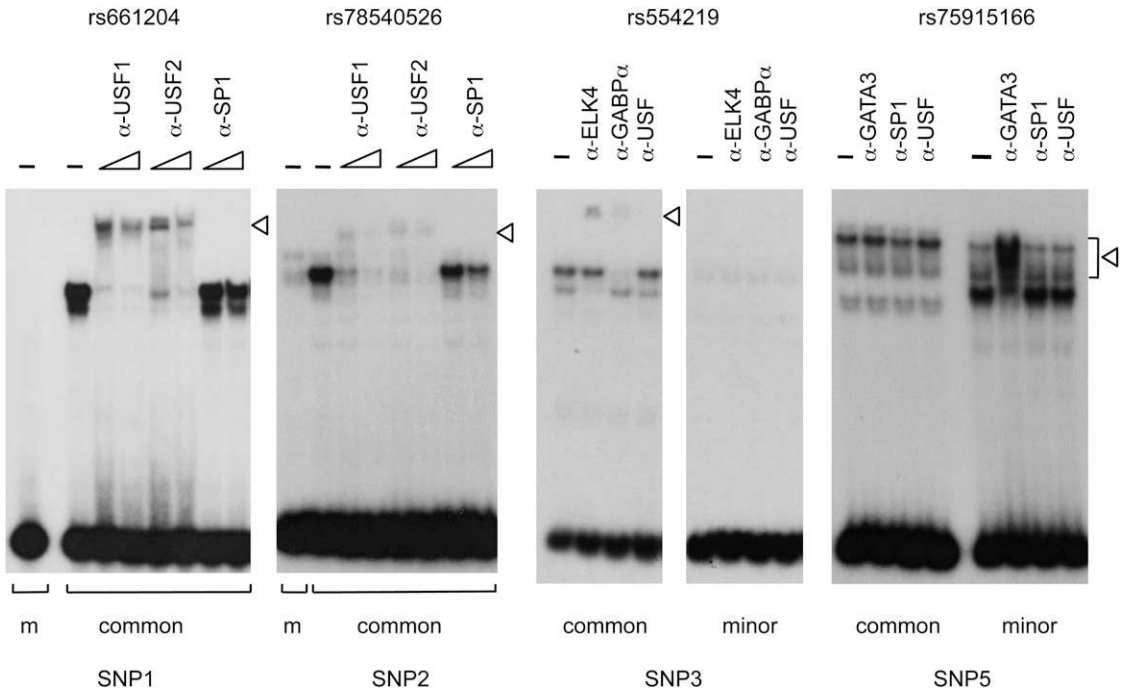

Figure 4. Allele-Specific In Vitro Protein-DNA Interactions Detected by EMSA
Nuclear extracts from MCF7 cells were incubated with radioactively labeled oligonucleotides overlapping the SNP shown at the top of each panel. The effect of the minor (m) and the common alleles are compared as indicated. 4 and 10 μg of antisera were included in the reaction as listed above each lane in panels 1 and 2, and 4 μg were used in all other reactions. Bands containing antibody-protein-DNA complexes are highlighted by open arrows.

common allele (Figure 5D). Thus, we conclude that in this context GATA3 acts as a repressor of transcription and that the risk alleles of both PRE1 SNP rs554219 (SNP3) and PRE2 SNP rs75915166 (SNP5) reduce transcriptional activation.

Discussion

Our fine-scale mapping of this 11q13 locus has identified three independent association signals, each with different effect sizes. We have been able to examine, in detail, the ones with the strongest effects. The hits are correlated with the originally detected GWAS tag SNP (rs614367; $r^2 = 0.87$ with rs554219 [SNP3], $r^2 = 0.31$ with rs78540526 [SNP2], and $r^2 = 0.57$ with rs75915166 [SNP5]). These strong candidates for being causative variants are more strongly associated with breast cancer than rs614367 (Table 1). In fact, the effect sizes of these newly recognized 11q13 SNPs are now larger than the effects of the best GWAS-discovered breast cancer locus, *FGFR2* (MIM 176943; OR overall breast cancer per minor allele = 1.31; 95% CI 1.26–1.36; p value = 2.93×10^{-44} for 11q13 rs75915166 versus 1.27; 95% CI 1.24–1.29; p value 10^{-129} for *FGFR2* rs2981579). Thus, by fine-scale mapping, we have also detected a little more of the "missing heritability" of breast cancer. On the basis of the estimates from this iCOGs study, the original GWAS tag SNP, rs614367, explains approximately 0.76% of the familial risk of overall breast cancer, whereas the combined effects of SNPs rs78540526 (SNP2), rs554219 (SNP3), and rs75915156 (SNP5) explain approximately 2.0% in Europeans.

Despite its clear value in this study, mapping by genetic epidemiological techniques alone, even in this very large BCAC consortium, was unable to differentiate three of the four candidates in signal 1 (SNPs 1, 3, and 4), which are very highly correlated in Europeans and rare in Asians, though we were able to demonstrate an independent effect for SNP2 (rs78540526). Of note, SNPs 3 and 4 (rs554219 and rs657686) are almost perfectly correlated ($r^2 = 0.998$) across all participating samples in the BCAC consortium. For signal 2, fine-scale mapping was more successful, because no other SNPs were strongly correlated with rs75915166 (SNP5). Our combined evidence suggests that SNP rs75915166 is functionally related to risk. However, it is important to bear in mind that when we selected mapping SNPs to go onto the iCOGs chip (in March 2010), the catalog of all common variants in the locus was not complete. Since then, reinterrogation of the 1000 Genomes data set and imputation of missing SNPs, with the most recent (January 2012) version, has indicated a new candidate (at chromosome 11 SNP rs12575120). It remains possible that other candidate causal variants may have been missed. It is worth noting, however, that the existence of three haplotypes associated with different risks makes it extremely unlikely (even in the absence of functional evidence) that the associations could be driven by rare variants missed by sequencing, because this would

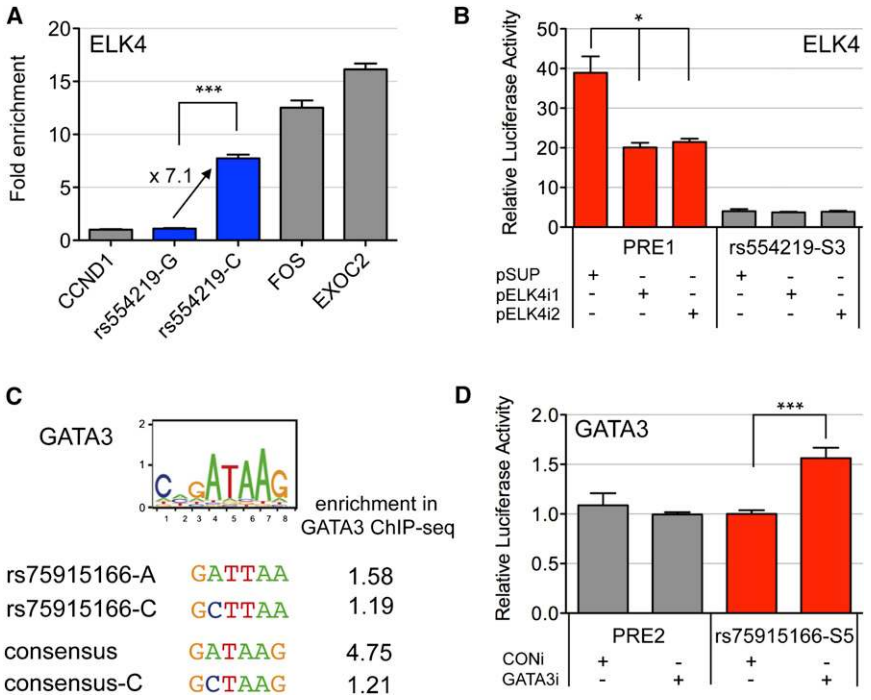

Figure 5. Allele-Specific Effects of ELK4 and GATA3 In Vivo

(A) ChIP assays by means of polyclonal ELK4 antiserum were carried out in MDAMB415 cells heterozygous (G/C) for rs554219. A TaqMan assay was used to detect allele-specific enrichment of rs554219, which is given relative to a negative control from the fourth intron of *CCND1*. Promoter sequences from *FOS* (MIM 164810) and *EXOC3* (MIM 608186) were used as positive controls.

(B) Luciferase reporter assays showing the effect of shRNA ELK4 silencing on the activity of PRE1 containing different alleles of rs554219. Error bars denote SD from two biological replicates performed in triplicate.

(C) Allele-specific changes to in vivo binding of GATA3. The position weight matrix of GATA3 derived in MCF7 cells is shown and compared to the sequence surrounding rs75915166. Fold enrichment of the A versus the C allele under GATA3 ChIP-seq peaks as compared to random genomic sequences is given for rs75915166 and the GATA3 consensus binding site.

(D) Luciferase reporter assays showing the effect of GATA3 siRNA silencing on the activity of PRE2 containing different alleles of rs75915166. Error bars denote SD from two biological replicates performed in triplicate. p values were determined with a two-tailed t test. *$p < 0.05$, ***$p < 0.001$. Red bars indicate constructs that contain either an ELK4 or GATA3 binding site. Levels of ELK4 and GATA3 repression are shown in Figure S10.

require at least two rare variants on different haplotypes conferring implausibly large effects.

We used functional studies to further examine the five best candidates. We have demonstrated that SNPs 1–4 all map to a putative *CCND1* enhancer (PRE1) and that the most plausible causal variant, SNP rs554219 (SNP3), alters binding of the ELK4 TF both in vitro and in vivo. The protective C allele of rs554219 preferably binds ELK4 and absence of binding at the minor allele strongly reduces enhancer activity in luciferase assays. This effect can be mimicked by transfection of ELK4 siRNA. Furthermore, presence of the risk allele correlates with reduced cyclin D1 protein levels. Evidence for a functional role for SNP rs78540526 (SNP2) is also strong: it is present on both haplotypes associated with greatest breast cancer risk, as well as significantly reducing enhancer activity in luciferase assays and also displaying allele-specific binding by TFs USF1 and USF2 in in vitro, but not in vivo, studies. Finally, we demonstrate that the effects of SNP rs75915166 (SNP5) are likely to be mediated via differential binding of TF GATA3 to this SNP position. SNP rs75915166 lies within a silencer element able to physically interact with the PRE1 enhancer containing SNPs 1–4. It has not yet been possible in this study to investigate the functions of the other potential risk variants. However, bioinformatic analysis suggests that the T allele of SNP rs494406 may form a GATA1 binding site and SNP rs585568 falls in a USF ChIP-seq peak, with the minor allele forming a MYC-MAX TF binding site. An understanding of the relevance of these will require even larger association studies and then further functional analyses.

Our data implicate ELK4 and GATA3 as mediators of risk for ER^+ disease, which is consistent with previous reports of the functions of these TFs. Expression of ELK4 is sensitive to ER inhibitors[23] and ChIP-seq data reveal a strong ER binding site upstream of the *ELK4* promoter. GATA3 has long been established as a critical regulator of mammary gland development and luminal epithelial differentiation,[24,25] and loss of GATA3 is associated with marked progression to early carcinoma.[26] At the molecular level, GATA3 influences ERα binding by modulating chromatin structure and long-range looping[17] and, along with FOXA1 and ERα, is a critical component of a cooperative network of transcriptional master regulators,[16,24,27] which are sufficient to confer estrogen responsiveness to ER-negative cell lines.[28] A GATA3 link with ER^+ breast cancer is further supported by the finding that tumors carrying *GATA3* mutations are all of a luminal subtype.[29,30]

Our findings support a hypothesis that *CCND1* is the target gene of these candidate causative SNPs—demonstrated by the strong physical interactions between the PREs at this locus and *CCND1* and by the fact that the risk alleles act by reducing the transcriptional activation of *CCND1*. In luciferase assays, the functional risk SNPs examined act to reduce transcriptional activity. These conclusions may be supported by our observation of reduced cyclin D1 protein levels in tumors homozygote for the G allele of rs554219, but we failed to detect similar

associations in RNA expression data. Our power to detect any such association was limited—we estimate that the 300 TCGA tumors analyzed provided 70% power to detect a 10% difference in expression associated with this risk allele (MAF = 0.12). However, it is also possible that any effect of this SNP on expression levels is not apparent in breast tumor cells. There is precedent for this in that the confirmed multicancer risk SNPs at 8q24 (upstream of *MYC*)[31] have consistently failed to show any associations with gene expression in human cell types but have been confirmed as functionally important in this respect when analyzed in transgenic mouse models.[32]

Cyclin D1 is traditionally considered to be an oncogene, based on its overexpression in tumors, its well-established role in cell cycle control, and its ability to promote cell migration and differentiation.[33,34] Consequently, germline variants that repress this gene are somewhat at odds with the accepted dogma of cancer-susceptibility genes. Resolution of this apparent conflict may, however, come from the complexity of cyclin D1 function, the heterogeneity of cyclin D1 protein levels in human tumors, and the fact that the moderate risk we describe is likely to work in concert with a number of other coinherited variants that may facilitate some lesser known activities of this protein.

In terms of function, repression of cyclin D1 has been reported to induce cell migration of breast cancer cell lines and be associated with the epithelial-mesenchymal transition (EMT).[35] Cyclin D1 also interacts with a range of TFs, including steroid hormone receptors[36,37] and chromatin-modifying enzymes,[37,38] and is able to participate in a broad range of other functions. A recent study provides evidence that cyclin D1 promotes homologous recombination-mediated DNA repair (HRR) by recruiting RAD51 to double-strand breaks, a role that is independent of its control of the cell cycle.[39] Notably, depletion of *CCND1* levels impairs HRR and increases sensitivity of cells to DNA-damaging agents such as ionizing radiation in vitro and in vivo.[39] It is thus conceivable that, in a similar way to BRCA1 and BRCA2 that also function in HRR, reduced cyclin D1 levels may lead to more error-prone repair mechanisms, potentially promoting genome instability and cancer predisposition.

Many of these roles are in fact more in-line with a tumor suppressor, suggesting that cyclin D1 can operate both as an oncogene and a tumor suppressor depending on the context, with the latter being particularly relevant in the case of germline events resulting in loss of cyclin D1. There are certainly precedents for this: *RET* (MIM 164761), for example, acts as an oncogene in the thyroid gland and a tumor-suppressor gene in the colon.[40] Consistent with this, Lehn and colleagues have reported an association between downregulation of cyclin D1 and unfavorable prognosis in human breast cancer.[41] We therefore propose that germline events leading to a reduction of cyclin D1, such as described in the manuscript, contribute to breast tumorigenesis.

Although our data indicate that *CCND1* is the target gene, we cannot rule out the possibility that these SNPs also exert functional effects through long-range control of other nearby genes: *MYEOV, ORAOV1*, or *FGF3*, *FGF4*, or *FGF19,* all of which are plausible candidates for breast cancer susceptibility. *MYEOV* is very frequently coamplified with *CCND1* in ER^+ breast cancer[42] and cyclin D1 protein levels are reduced in HeLa cells in which *ORAOV1* proto-oncogene levels have been knocked down to induce apoptosis;[43] *FGF3*, *FGF4*, and *FGF19* belong to the Fibroblast Growth Factor family, which plays a key role in tumor pathogenesis via their receptors—including *FGFR2*, the strongest common breast cancer susceptibility locus reported to date.[44–46]

This 11q13 genomic interval also contains GWAS hits for renal cancer and functional studies there indicate that a candidate causal variant affects the activity of an enhancer element probably also driving *CCND1* transcription.[47] This region may thus share similarities with the 8q24 interval, where a large "gene desert" contains multiple tissue-specific enhancers in which risk-associated SNPs affect the transcription of downstream oncogene(s) and predispose to cancer in a tissue-specific manner.[47] The variants described here predispose only to ER^+ disease and we find that the identified molecular mechanisms underlying this risk are fully consistent with this observation: long-range physical interactions between the enhancer and the *CCND1* promoter are found only in ER^+ cells and GATA3, which binds PRE1 SNP3 (rs554219), is coregulated with the ER and is part of the network of transcriptional master regulators able to establish estrogen responsiveness.[28,29,48]

In conclusion we have identified three SNPs as being very strong candidates for having a directly causative effect on breast cancer risk at this locus and we have provided evidence that these act by controlling *CCND1* expression—a gene that is a potential target for drug intervention.

Supplemental Data

Supplemental Data include Acknowledgments, ten figures, and four tables and can be found with this article online at http://www.cell.com/AJHG/.

Received: November 12, 2012
Revised: December 21, 2012
Accepted: January 3, 2013
Published: March 28, 2013

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://browser.1000genomes.org/index.html
ArrayExpress, http://www.ebi.ac.uk/arrayexpress/
ChIA-PET Browser, http://cms1.gis.a-star.edu.sg/index.php
Gene Expression Omnibus (GEO), http://www.ncbi.nlm.nih.gov/geo/

National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/

NURSA Transcriptomine, http://www.nursa.org/transcriptomine

Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/

References

1. Lambrechts, D., Truong, T., Justenhoven, C., Humphreys, M.K., Wang, J., Hopper, J.L., Dite, G.S., Apicella, C., Southey, M.C., Schmidt, M.K., et al.; GENICA Network; Investigators kConFab; Australian Ovarian Cancer Study Group. (2012). 11q13 is a susceptibility locus for hormone receptor positive breast cancer. Hum. Mutat. *33*, 1123–1132.
2. 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.
3. Purdue, M.P., Johansson, M., Zelenika, D., Toro, J.R., Scelo, G., Moore, L.E., Prokhortchouk, E., Wu, X., Kiemeney, L.A., Gaborieau, V., et al. (2011). Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. Nat. Genet. *43*, 60–65.
4. Eeles, R.A., Kote-Jarai, Z., Giles, G.G., Olama, A.A., Guy, M., Jugurnauth, S.K., Mulholland, S., Leongamornlert, D.A., Edwards, S.M., Morrison, J., et al.; UK Genetic Prostate Cancer Study Collaborators; British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. Nat. Genet. *40*, 316–321.
5. Gudmundsson, J., Sulem, P., Gudbjartsson, D.F., Blondal, T., Gylfason, A., Agnarsson, B.A., Benediktsdottir, K.R., Magnusdottir, D.N., Orlygsdottir, G., Jakobsdottir, M., et al. (2009). Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nat. Genet. *41*, 1122–1126.
6. Schumacher, F.R., Berndt, S.I., Siddiq, A., Jacobs, K.B., Wang, Z., Lindstrom, S., Stevens, V.L., Chen, C., Mondul, A.M., Travis, R.C., et al. (2011). Genome-wide association study identifies new prostate cancer susceptibility loci. Hum. Mol. Genet. *20*, 3867–3875.
7. Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A., et al. (2008). Multiple loci identified in a genome-wide association study of prostate cancer. Nat. Genet. *40*, 310–315.
8. 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. (2013). Large-scale genotyping identifies new loci associated with breast cancer risk. Nat. Genet. Published online March 27, 2013. http://dx.doi.org/10.1038/ng.2563.
9. Stram, D.O., Haiman, C.A., Hirschhorn, J.N., Altshuler, D., Kolonel, L.N., Henderson, B.E., and Pike, M.C. (2003). Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. Hum. Hered. *55*, 27–36.
10. Schaid, D.J., Rowland, C.M., Tines, D.E., Jacobson, R.M., and Poland, G.A. (2002). Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am. J. Hum. Genet. *70*, 425–434.
11. Aaltonen, K., Amini, R.M., Heikkilä, P., Aittomäki, K., Tamminen, A., Nevanlinna, H., and Blomqvist, C. (2009). High cyclin B1 expression is associated with poor survival in breast cancer. Br. J. Cancer *100*, 1055–1060.
12. Tan-Wong, S.M., French, J.D., Proudfoot, N.J., and Brown, M.A. (2008). Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. Proc. Natl. Acad. Sci. USA *105*, 5160–5165.
13. Day, B.W., Stringer, B.W., Spanevello, M.D., Charmsaz, S., Jamieson, P.R., Ensbey, K.S., Carter, J.C., Cox, J.M., Ellis, V.J., Brown, C.L., et al. (2011). ELK4 neutralization sensitizes glioblastoma to apoptosis through downregulation of the anti-apoptotic protein Mcl-1. Neuro-oncol. *13*, 1202–1212.
14. Meyer, K.B., Maia, A.T., O'Reilly, M., Teschendorff, A.E., Chin, S.F., Caldas, C., and Ponder, B.A. (2008). Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. PLoS Biol. *6*, e108.
15. Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J., and Odom, D.T. (2009). ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. Methods *48*, 240–248.
16. Eeckhoute, J., Carroll, J.S., Geistlinger, T.R., Torres-Arzayus, M.I., and Brown, M. (2006). A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer. Genes Dev. *20*, 2513–2526.
17. Theodorou, V., Stark, R., Menon, S., and Carroll, J.S. (2012). GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. Genome Res. *23*, 12–22.
18. Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell *132*, 958–970.
19. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.
20. Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58–64.
21. Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.
22. Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoute, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F., et al. (2006). Genome-wide analysis of estrogen receptor binding sites. Nat. Genet. *38*, 1289–1297.
23. Ochsner, S.A., Watkins, C.M., McOwiti, A., Xu, X., Darlington, Y.F., Dehart, M.D., Cooney, A.J., Steffen, D.L., Becnel, L.B., and McKenna, N.J. (2012). Transcriptomine, a web resource for nuclear receptor signaling transcriptomes. Physiol. Genomics *44*, 853–863.
24. Kouros-Mehr, H., Slorach, E.M., Sternlicht, M.D., and Werb, Z. (2006). GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. Cell *127*, 1041–1055.
25. Asselin-Labat, M.L., Sutherland, K.D., Barker, H., Thomas, R., Shackleton, M., Forrest, N.C., Hartley, L., Robb, L., Grosveld, F.G., van der Wees, J., et al. (2007). Gata-3 is an essential

regulator of mammary-gland morphogenesis and luminal-cell differentiation. Nat. Cell Biol. *9*, 201–209.
26. Kouros-Mehr, H., Bechis, S.K., Slorach, E.M., Littlepage, L.E., Egeblad, M., Ewald, A.J., Pai, S.Y., Ho, I.C., and Werb, Z. (2008). GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. Cancer Cell *13*, 141–152.
27. Eeckhoute, J., Keeton, E.K., Lupien, M., Krum, S.A., Carroll, J.S., and Brown, M. (2007). Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. Cancer Res. *67*, 6477–6483.
28. Kong, S.L., Li, G., Loh, S.L., Sung, W.K., and Liu, E.T. (2011). Cellular reprogramming by the conjoint action of ERα, FOXA1, and GATA3 to a ligand-inducible growth state. Mol. Syst. Biol. *7*, 526.
29. Usary, J., Llaca, V., Karaca, G., Presswala, S., Karaca, M., He, X., Langerød, A., Kåresen, R., Oh, D.S., Dressler, L.G., et al. (2004). Mutation of GATA3 in human breast tumors. Oncogene *23*, 7669–7678.
30. Arnold, J.M., Choong, D.Y., Thompson, E.R., Waddell, N., Lindeman, G.J., Visvader, J.E., Campbell, I.G., and Chenevix-Trench, G.; kConFab. (2010). Frequent somatic mutations of GATA3 in non-BRCA1/BRCA2 familial breast tumors, but not in BRCA1-, BRCA2- or sporadic breast tumors. Breast Cancer Res. Treat. *119*, 491–496.
31. Ghoussaini, M., Song, H., Koessler, T., Al Olama, A.A., Kote-Jarai, Z., Driver, K.E., Pooley, K.A., Ramus, S.J., Kjaer, S.K., Hogdall, E., et al.; UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators. (2008). Multiple loci with different cancer specificities within the 8q24 gene desert. J. Natl. Cancer Inst. *100*, 962–966.
32. Sur, I.K., Hallikas, O., Vähärautio, A., Yan, J., Turunen, M., Enge, M., Taipale, M., Karhu, A., Aaltonen, L.A., and Taipale, J. (2012). Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. Science *338*, 1360–1363.
33. Ormandy, C.J., Musgrove, E.A., Hui, R., Daly, R.J., and Sutherland, R.L. (2003). Cyclin D1, EMS1 and 11q13 amplification in breast cancer. Breast Cancer Res. Treat. *78*, 323–335.
34. Zhong, Z., Yeow, W.S., Zou, C., Wassell, R., Wang, C., Pestell, R.G., Quong, J.N., and Quong, A.A. (2010). Cyclin D1/cyclin-dependent kinase 4 interacts with filamin A and affects the migration and invasion potential of breast cancer cells. Cancer Res. *70*, 2105–2114.
35. Tobin, N.P., Sims, A.H., Lundgren, K.L., Lehn, S., and Landberg, G. (2011). Cyclin D1, Id1 and EMT in breast cancer. BMC Cancer *11*, 417.
36. Zwijsen, R.M., Buckle, R.S., Hijmans, E.M., Loomans, C.J., and Bernards, R. (1998). Ligand-independent recruitment of steroid receptor coactivators to estrogen receptor by cyclin D1. Genes Dev. *12*, 3488–3498.
37. Fu, M., Rao, M., Bouras, T., Wang, C., Wu, K., Zhang, X., Li, Z., Yao, T.P., and Pestell, R.G. (2005). Cyclin D1 inhibits peroxisome proliferator-activated receptor gamma-mediated adipogenesis through histone deacetylase recruitment. J. Biol. Chem. *280*, 16934–16941.
38. Bienvenu, F., Jirawatnotai, S., Elias, J.E., Meyer, C.A., Mizeracka, K., Marson, A., Frampton, G.M., Cole, M.F., Odom, D.T., Odajima, J., et al. (2010). Transcriptional role of cyclin D1 in development revealed by a genetic-proteomic screen. Nature *463*, 374–378.
39. Jirawatnotai, S., Hu, Y., Michowski, W., Elias, J.E., Becks, L., Bienvenu, F., Zagozdzon, A., Goswami, T., Wang, Y.E., Clark, A.B., et al. (2011). A function for cyclin D1 in DNA repair uncovered by protein interactome analyses in human cancers. Nature *474*, 230–234.
40. Luo, Y., Tsuchiya, K.D., Il Park, D., Fausel, R., Kanngurn, S., Welcsh, P., Dzieciatkowski, S., Wang, J., and Grady, W.M. (2012). RET is a potential tumor suppressor gene in colorectal cancer. Oncogene. Published online July 2, 2012. http://dx.doi.org/10.1038/onc.2012.225.
41. Lehn, S., Tobin, N.P., Berglund, P., Nilsson, K., Sims, A.H., Jirström, K., Härkönen, P., Lamb, R., and Landberg, G. (2010). Down-regulation of the oncogene cyclin D1 increases migratory capacity in breast cancer and is linked to unfavorable prognostic features. Am. J. Pathol. *177*, 2886–2897.
42. Janssen, J.W., Cuny, M., Orsetti, B., Rodriguez, C., Vallés, H., Bartram, C.R., Schuuring, E., and Theillet, C. (2002). MYEOV: a candidate gene for DNA amplification events occurring centromeric to CCND1 in breast cancer. Int. J. Cancer *102*, 608–614.
43. Jiang, L., Zeng, X., Wang, Z., Ji, N., Zhou, Y., Liu, X., and Chen, Q. (2010). Oral cancer overexpressed 1 (ORAOV1) regulates cell cycle and apoptosis in cervical cancer HeLa cells. Mol. Cancer *9*, 20.
44. Adnane, J., Gaudray, P., Dionne, C.A., Crumley, G., Jaye, M., Schlessinger, J., Jeanteur, P., Birnbaum, D., and Theillet, C. (1991). BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. Oncogene *6*, 659–663.
45. Jang, J.H., Shin, K.H., and Park, J.G. (2001). Mutations in fibroblast growth factor receptor 2 and fibroblast growth factor receptor 3 genes associated with human gastric and colorectal cancers. Cancer Res. *61*, 3541–3543.
46. Moffa, A.B., Tannheimer, S.L., and Ethier, S.P. (2004). Transforming potential of alternatively spliced variants of fibroblast growth factor receptor 2 in human mammary epithelial cells. Mol. Cancer Res. *2*, 643–652.
47. Schödel, J., Bardella, C., Sciesielski, L.K., Brown, J.M., Pugh, C.W., Buckle, V., Tomlinson, I.P., Ratcliffe, P.J., and Mole, D.R. (2012). Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. Nat. Genet. *44*, 420–425, S1–S2.
48. Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.