Functional annotation of the 2q35 breast cancer risk locus implicates a structural variant in influencing activity of a long-range enhancer element

Joseph S. Baxter,1,* Nichola Johnson,1 Katarzyna Tomczyk,1 Andrea Gillespie,1 Sarah Maguire,2 Rachel Brough,1,3 Laura Fachal,4 Kyriaki Michailidou,5,6,7 Manjeet K. Bolla,7 Qin Wang,7 Joe Dennis,7 Thomas U. Ahearn,8 Irene L. Andrusik,9,10 Hoda Anton-Culver,11 Natalia N. Antonenkova,12 Volker Arndt,13 Kristan J. Aronson,14 Annelie Augustinsson,15 Heiko Becher,16 Matthias W. Beckmann,17 Sabine Behrens,18 Javier Benitez,19,20 Marina Bermisheva,21 Natalia V. Bogdanova,12,22,23 Stig E. Bojesen,24,25,26 Hermann Brenner,13,27,28 Sara Y. Brucker,29 Quyin Cai,30 Daniele Campa,18,31 Federico Canzian,32 Jose E. Castelao,33 Tsun L. Chan,34,35 Jenny Chang-Claude,18,36 Stephen J. Chanock,8 Georgia Chenevix-Trench,37 Ji-Yeob Choi,38,39,40 Christine L. Clarke,41 NBCS Collaborators,42,43,44,45,46,47,48,49,50,51,52 Sarah Colonna,53 Don M. Conroy,4 Fergus J. Couch,54 Angela Cox,55 Simon S. Cross,56 Kamila Czene,57 Mary B. Daly,58 Peter Devilee,59,60 Thilo Dörk,23 Laure Dossus,61 Miriam Dwek,62 Diana M. Eccles,63 Arif B. Ekici,64 A. Heather Eliassen,65,66

(Author list continued on next page)

Abstract

A combination of genetic and functional approaches has identified three independent breast cancer risk loci at 2q35. A recent fine-scale mapping analysis to refine these associations resulted in 1 (signal 1), 5 (signal 2), and 42 (signal 3) credible causal variants at these loci. We used publicly available in silico DNase I and ChIP-seq data with in vitro reporter gene and CRISPR assays to annotate signals 2 and 3. We identified putative regulatory elements that enhanced cell-type-specific transcription from the IGFBP5 promoter at both signals (30- to 40-fold increased expression by the putative regulatory element at signal 2, 2- to 3-fold by the putative regulatory element at signal 3). We further identified one of the five credible causal variants at signal 2, a 1.4 kb deletion (esv3594306), as the likely causal variant; the deletion allele of this variant was associated with an average additional increase in IGFBP5 expression of 1.3-fold (MCF-7) and 2.2-fold (T-47D). We propose a model in which the deletion allele of esv3594306 juxtaposes two transcription factor binding regions (annotated by estrogen receptor alpha ChIP-seq peaks) to generate a single extended regulatory element. This regulatory element increases cell-type-specific expression of the tumor suppressor gene IGFBP5 and, thereby, reduces risk of estrogen receptor-positive breast cancer (odds ratio = 0.77, 95% CI 0.74–0.81, p = 3.1 × 10–31).

Introduction

Over the last 15 years, genome-wide association studies have transformed our ability to map genetic variation underpinning complex traits.1 The vast majority of variants identified in genome-wide association studies are non-coding and are thought to influence transcriptional regulation,2,3 a process which can be highly cell type and tissue

1The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London SW7 3RP, UK; 2Centre for Cancer Research and Cell Biology, Queen’s University Belfast, Belfast, Ireland BT7 1NN, UK; 3The CRUK Gene Function Laboratory, The Institute of Cancer Research, London SW3 6JB, UK; 4Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK; 5Biostatistics Unit, The Cyprus Institute of Neurology & Genetics, Nicosia 2171, Cyprus; 6Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology & Genetics, Nicosia 2171, Cyprus; 7Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK; 8Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20850, USA; 9Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON M5G 1XS, Canada; 10Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada; 11Department of Medicine, Genetic Epidemiology Research Institute, University of California Irvine, Irvine, CA 92617, USA; 12N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk 223040, Belarus; 13Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany; 14Department of Public Health Sciences, and Cancer Research Institute, Queen’s University, Kingston, ON K7L 3N6, Canada; 15Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund 222 42, Sweden; 16Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany; 17Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg (FAU), Erlangen 91054, Germany; 18Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany; 19Biomedical Network on Rare Diseases (CIBERER), Madrid 28029, Spain; 20Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid 28029, Spain; 21Institute of Biochemistry and Genetics, Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa 450054, Russia; 22Department of Radiation Oncology, Hannover Medical School, Hannover 30625, Germany; 23Gynaecology Research Unit, Hannover Medical School, Hannover 30625, Germany; 24Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev 2730, Denmark; 25Department of Clinical Biochemistry, Herlev and
University of Singapore, Singapore 119077, Singapore; 90Department of Pathology, The Ohio State University, Columbus 43210, USA; 91Department of Clinical and Translational Health Sciences, Mayo Clinic, Rochester, MN 55905, USA; 92Department of Medical Oncology, Karolinska Institutet, Stockholm 171 77, Sweden; 93Department of Surgery, National University Hospital, Singapore 119228, Singapore; 94Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119077, Singapore; 95Laboratory of Experimental Oncology (LEO), Department of Oncology, KU Leuven, Leuven Cancer Institute, Leuven 3000, Belgium; 96Center for Molecular Medicine Cologne (CMMC), Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne 50931, Germany; 97Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam 3015 GD, the Netherlands; 98Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart 70376, Germany; 99University of Tubingen, Tubingen 72074, Germany; 100Department of Surgery, Kaohsiung Municipal Hsiao-Kang Hospital, Kaohsiung 812, Taiwan; 101Research Department, Peter MacCallum Cancer Center, Melbourne, VIC 3000, Australia; 102Sir Peter MacCallum Department of Oncology, The University of Melbourne, Melbourne, VIC 3000, Australia; 103Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, NSW 2145, Australia; 104Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya 464-8681, Japan; 105Division of Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan; 106Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, Tokyo 104-0045, Japan; 107Department of Genetics and Pathology, Pomeranian Medical University, Szczecin 71-252, Poland; 108Independent Laboratory of Molecular Biology and Genetic Diagnostics, Pomeranian Medical University, Szczecin 71-252, Poland; 109Department of Gynecology and Obstetrics, University Hospital Ulm, Ulm 89075, Germany; 110Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA 94305, USA; 111Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94304, USA; 112Clinical Genetics Research Lab, Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA; 113Department of Preventive Medicine, National University of Singapore, College of Medicine, Seoul 03080, Korea; 114Division of Molecular Pathology, the Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam 1066 CX, the Netherlands; 115Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa 450000, Russia; 116Department of Surgery, Daerim Saint Mary's Hospital, Seoul 07442, Korea; 117Translational Cancer Research Area, University of Eastern Finland, Kuopio 70210, Finland; 118Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio 70210, Finland; 119Biobank of Eastern Finland, Kuopio University Hospital, Kuopio, Finland; 120Department of Medical Genetics, Osaka University Hospital and University of Osaka, Osaka 0379, Norway; 121Department of Histopathology and Cytology, Clinical Hospital Acidem Stintina, Skokie 1000, Republic of North Macedonia; 122Department of Surgery, The University of Hong Kong, Hong Kong; 123Department of Surgery and Cancer Genetics Center, Hong Kong Sanatorium and Hospital, Hong Kong; 124Department of Computational and Quantitative Medicine, City of Hope, Duarte, CA 91010, USA; 125City of Hope Comprehensive Cancer Center, City of Hope, Duarte, CA 91010, USA; 126VIB Center for Cancer Biology, Leuven 3001, Belgium; 127Laboratory for Translational Genetics, Department of Human Genetics, University of Leuven, Leuven 3000, Belgium; 128Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; 129Institute of Environmental Medicine, Karolinska Institutet, Stockholm 171 77, Sweden; 130Department of Surgical Sciences, Uppsala University, Uppsala 751 05, Sweden; 131Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI 96813, USA; 132Clift National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa 35254, Israel; 133Human Genetics Division, Genome Institute of Singapore, Singapore 138672, Singapore; 134Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester M13 9PL, UK; 135Unit of Medical Genetics, Department of Medical Oncology and Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan 20133, Italy; 136Department of Clinical Science and Education, Sodersjukhuset, Karolinska Institutet, Stockholm 118 83, Sweden; 137Department of Medical Oncology, University Hospital of Heraklion, Heraklion 711 10, Greece; 138Institute of Clinical Trials & Methodology, University College London, London WC1V 6LJ, UK; 139Breast Cancer Research Unit, University Malaya Cancer Research Institute, Faculty of Medicine, University of Malaya, Kuala Lumpur 50603, Malaysia; 140Department of Obstetrics and Gynecology, Helsinki University Hospital, Helsinki 00290, Finland; 141School of Population and Public Health, University of British Columbia, Vancouver, BC V6T 1Z4, Canada; 142Cancer Control Research, BC Cancer, Vancouver, BC VSZ 1L3, Canada; 143Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC 27709, USA; 144Clinical Genetics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA; 145Convergence Graduate Program in Innovative Medical Science, Seoul National University College of Medicine, Seoul 03080, Korea; 146Department of Population Science, American Cancer Society, Atlanta, GA 30303, USA; 147Genome Diagnostics Program, FOM - the FIRC Institute of Molecular Oncology, Milan 20139, Italy; 148Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK; 149Research Centre for Genetic Engineering and Biotechnology ‘Georgi D. Efremov’, NASA, Skopo 1000, Republic of North Macedonia; 150Laboratory of Cancer Genetics and Tumor Biology, Cancer and Translational Medicine Research Unit, Biocenter Oulu, University of Oulu, Oulu 90570, Finland; 151Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre Oulu, Oulu 90570, Finland; 152Medical Oncology Department, Hospital Universitario Puerta de Hierro, Madrid 28222, Spain; 153Institute of Pathology, Staedtisches Klinikum Karlsruhe, Karlsruhe 76133, Germany; 154Division of Oncology, University Hospital of Larissa, Larissa 411 10, Greece; 155School of Cancer & Pharmaceutical Sciences, Comprehensive Cancer Centre, Guy's Campus, King's College London, London, UK; 156Division of Psychosocial Research and Epidemiology, the Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam 1066 CX, the Netherlands; 157National Center for Tumor Diseases, University of Göttingen and German Cancer Research Center, Heidelberg 69120, Germany; 158Division of Genetics and Epidemiology, The Institute of Cancer Research, London SM2 5NG, UK; 159Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan; 160School of Public Health, China Medical University, Taichung, Taiwan; 161Genomics Center, Centre Hospitalier Universitaire de Quebec - Universite Laval Research Center, Quebec City, QC G1V 4G2, Canada; 162Department of Clinical Pathology, The University of Melbourne, Melbourne, VIC 3010, Australia; 163Genetic Epidemiology Group, School of Population and Global Health, University of Western Australia, Perth, WA 6000, Australia; 164Division of Breast Cancer Research, The Institute of Cancer Research, London SW7 3RP, UK; 165Department of Non-Cancer Health Sciences, Weill Cornell Medicine, New York, NY 10065, USA; 166Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research
specific.4 Our ability to translate these findings into a greater understanding of the mechanisms that influence an individual woman’s risk will require the identification of causal variants (as opposed to correlated variants), the targets of these functional variants (the genes or non-coding RNAs that mediate the associations observed in genome-wide association studies) and an understanding of the disease causal cell types and processes.4 Genome-wide association studies of breast cancer coupled with large-scale replication and fine-mapping studies have led to the identification of approximately 200 breast cancer risk loci,5–9 two of these loci, annotated by rs13387042 and rs16857609,5 map to a gene desert at chromosome 2q35. Fine-scale mapping, combined with in silico annotation, reporter gene assays, and allele-specific qRT-PCR led to the identification of a putative causal variant (rs4442975) at the rs13387042 locus.11,12 rs4442975, which is highly correlated with the tag SNP rs13387042 (r² = 0.92, D’ = 0.96), maps to a consensus binding site for the transcription factor (TF) forkhead box A1 (FOXA1 [MIM: 602294]) with the alternative T-allele promoting binding of FOXA1.11,12 To date, no putative causal variant at the rs16857609 locus has been reported. Chromatin interaction methods implicate IGFBP5 (MIM: 146734) as the target gene at both loci11–13 and for the rs13387042 locus, eQTL analyses demonstrated association of the protective T-allele with slightly increased IGFBP5 levels in normal breast tissue11 and estrogen receptor-positive (ER+) breast cancers.12

Taking a functional approach based on chromosome conformation capture (3C) assays that were anchored at the IGFBP5 promoter, Wyszynski and colleagues identified a putative regulatory element centered on a structural variant (SV; esv3594306) that maps approximately 400 kb telomeric to IGFBP5.14 Allele-specific expression analyses and follow-up genotyping identified 14 highly correlated variants (all r² > 0.8 with the top SNP, rs34005590) associated with breast cancer risk, which represent a third risk signal (OR = 0.82, p = 5.6 × 10⁻¹⁷).14

In this analysis we report fine-scale mapping of the 2q35 region in European and Asian individuals with breast cancer and control subjects from the Breast Cancer Association Consortium. We confirm three independent, high-confidence signals at 2q35 annotated by rs13387042 (signal 1), rs138522813 (signal 2), and rs16857609 (signal 3). We carry out functional annotation of credible variants at signals 2 and 3 and implicate the deletion variant (esv3594306) at signal 2 as causally associated with increased IGFBP5 expression and reduced breast cancer risk.

**Material and methods**

**Fine-scale mapping of the 2q35 breast cancer risk locus**

Fine-scale mapping of the 2q35 breast cancer risk locus was carried out as part of a large collaborative project; full details have been published.3 Briefly, for the current analysis we accessed data from 94,391 individuals with invasive breast cancer and 83,477 individuals of European ancestry and 12,481 individuals with invasive breast cancer and 12,758 control subjects of Asian ancestry from 87 studies participating in the Breast Cancer Association Consortium. All participating studies were approved by their appropriate ethics review board and all subjects provided informed consent.

Directly genotyped or imputed (info score > 0.8) calls for 10,314 SNPs mapping to a 1.4 Mb region at 2q35 (chr2:217,405,832–218,796,508; GRCh37/hg19) were available for analysis. At this threshold, the proportions of common variants (MAF ≥ 0.05), low-frequency variants (0.01 ≤ MAF < 0.05), and rare variants (0.001 ≤ MAF < 0.01)4 that could be analyzed were 89.7%, 68.5%, and 3.6%, respectively, for OncoArray and 64.2%, 40.5%, and 0.8%, respectively, for iCOGS. Analysis of the association between each SNP and risk of breast cancer was performed using unconditional logistic regression assuming a log-additive genetic model, adjusted for study and up to 15 ancestry-informative principal components. p values were calculated using Wald tests. Forward stepwise logistic regression was used to explore whether additional loci in the fine-mapping region were independently associated with breast cancer risk. We carried out stratified analyses to determine whether each of the independent associations...
differed according to estrogen receptor (ER) status; heterogeneity between stratum-specific estimates was assessed using Cochran’s Q-test. All statistical analyses were carried out using R version 3.6.1.

In silico annotation of credible variants

Credible variants at each of the three independent signals were aligned with DNase I and ChiP-seq data (P300 [EP300 (MIM: 602700)], H3K27Ac, H3K4me1, FOXA1, GATA3 [MIM: 131320], Erz [ESR1 (MIM: 133430)]) generated in T-47D and MCF-7 breast cancer cells$^{15-17}$ (Table S1).

Cloning of reporter assay constructs

All reporter assay plasmids were derived using the pGL4 reporter vector (Promega). Reporter vectors were constructed using a restriction digest-based cloning approach. The IGFBP5 promoter and putative regulatory element regions (containing WT alleles) were synthesized as gblocks (Integrated DNA Technologies, full details in Table S2). Double-restriction digest of plasmid or gBlock were performed using BgIII and XhoI (for IGFBP5 promoter) or Sall and BamHII (for putative regulatory element regions) according to the manufacturer’s instructions (New England Biolabs [NEB]). Ligations were performed in a 3:1 insert:vector ratio using T4 DNA ligase (NEB), according to manufacturer’s instructions. Correct cloning was validated by Sanger sequencing using a commercially available service (Eurofins Genomics). Alternative (ALT) alleles of each variant were introduced into reporter vectors using QuickChange Lightning Site-directed Mutagenesis kit (Agilent Technologies), according to the manufacturer’s instructions. Accurate mutagenesis was confirmed by Sanger sequencing (Eurofins Genomics). All reporter gene constructs are shown in Figure S1.

Cell Culture

T-47D cells were grown in RPMI (Gibco) supplemented with 10% FBS (GIBCO), 10 μg/mL human insulin (Sigma), and 100 μM penicillin with 100 μg/ml streptomycin (Sigma). HCT116 cells were grown in RPMI supplemented with 10% FBS, 100 μM penicillin, and 100 μg/mL streptomycin. HepG2 cells were grown in EMEM (LGC Standards-ATCC) supplemented with 10% FBS and 100 μM penicillin with 100 μg/mL streptomycin. MCF-7 cells (including derivative Cas9-expressing cell lines) and 293T cells were grown in DMEM (Gibco) supplemented with 10% FBS and 100 μM penicillin with 100 μg/mL streptomycin. All cell lines were routinely short tandem repeat (STR)-typed and tested for mycoplasma contamination.

Reporter assays

Reporter assays were performed in T-47D, MCF-7, 293T, HCT116, and HepG2 cell lines. Antibiotics were removed from standard growth media 24 h before transfection to improve viability. For assays performed under standard conditions, approximately 16,000 cells were seeded per well of a 96-well plate for T-47D, MCF-7, and HepG2, and approximately 8,000 cells were seeded per well of a 96-well plate for 293T and HCT116. Transfection was performed upon reaching 70% confluency (~24 h after cell seeding). For assays performed after 17β-estradiol treatment, cells were first hormone starved for 48 h. Approximately 10,000 cells (T-47D) and 8,000 cells (MCF-7) were seeded, per well of a 96-well plate, in standard growth media and cultured for 24 h. The media was then replaced with phenol red-free media (Gibco) supplemented with 10% charcoal-stripped FBS (GIBCO), 100 μM penicillin with 100 μg/mL streptomycin, 10 nM fulvestrant (I4409, Sigma), and 10 μg/mL human insulin (T-47D only). After 48 h, growth media was replaced with phenol red-free media supplemented with 10% charcoal-stripped FBS, 10 μg/mL human insulin (T-47D only), with the addition of either (1) 10 nM 17β-estradiol (E2758, Sigma) or (2) vehicle (ethanol). Transfection was performed upon reaching 80% confluency (6 h after 17β-estradiol or vehicle treatment).

Transfection was performed using X-treme GENE HP DNA transfection reagent (Roche). Equimolar amounts of the test pGL4-based firefly luciferase vector and pRL-TK renilla luciferase control (Promega) were combined in a 1:1 reagent:DNA ratio in OptiMEM (Fisher Scientific). After a 30 min incubation at room temperature, 10 μL transfection mixture was added per well. Each biological replicate was performed in technical triplicates with non-transfected, mock-transfected, and pEGFP-transfected controls (Takara Bio Inc). Cells were screened for luciferase activity 48 h after transfection using the Dual-Glo Luciferase Assay System (Promega) according to the manufacturer’s instructions.

Confirmatory genotyping and sequencing of putative regulatory element 2 (PRE2)

Four of the five variants mapping to PRE2 (rs72951831, rs199804270, rs138522813, and esv3594306) are highly correlated based on 1000 Genomes data (1KGP), with the ALT alleles of rs72951831, rs199804270, and rs138522813 all predicted to occur in combination with the ALT (deletion) allele of esv3594306 (rs3594306: rs72951831 $r^2$ = 1.0, $D^\prime$ = 1.0; esv3594306: rs199804270 $r^2$ = 0.95, $D^\prime$ = 1.0; esv3594306: rs138522813 $r^2$ = 1.0, $D^\prime$ = 1.0) . However, rs572022984 (hg19, chr2:217955897) theoretically maps within the esv3594306 deleted region (chr2:217,955,891-217,957,273) casting doubt on whether the (imputed) rs572022984-del allele could occur in combination with the esv3594306 deletion allele. To clarify this, we genotyped all five variants in 300 randomly selected women participating in the Generations Study$^{18}$ using MassARRAY (Agena Bioscience; full details of primers available on request). The number of carriers of the alternative (A>-) allele at rs572022984 (MAF = 0.035) was 0 (expected number = 21; $p = 0.00002$). To confirm our genotyping, we carried out Sanger sequencing (Eurofins) of a 2.4 kb region spanning (chr2:217,955,866-217,958,000) in two individuals who were heterozygous at the linked PRE2 SNP rs138522813. Primers were: forward 5′-CCGTTCCCTCTCATGTCCTG-3′ and, reverse 5′-CTCTTCAGGCCAAGTCACAG-3′. Sequencing confirmed the presence of REF and ALT alleles of esv3594306, rs72951831, and rs199804270 (rs138522813 maps just outside the amplified region) but only REF alleles at rs572022984; on this basis we excluded rs572022984 from further analyses.

Cloning of guides for CRISPR-based enhancer perturbation

Guides were designed using the online design tool CHOPCHOP (http://chopchop.cbu.uib.no). Guides were selected based on their proximity to variants of interest and specificity scores. Full details are provided in Table S3. Cloning was performed essentially as described in Ran et al.$^{19}$ Briefly, guides were produced as two complementary oligonucleotides with overhangs to facilitate cloning. Oligos were annealed with T4 Polynucleotide Kinase (NEB). The expression vector pKLV-U6gRNA(BbsI)-PGKpuro2ABFP (Addgene #50946) was digested using BbsI (NEB), and ligation performed using T4 DNA ligase (NEB). Cloning was validated by sequencing (Eurofins Genomics).
CRISPR-based enhancer perturbation

All CRISPR cell lines were derived from a parental MCF-7 cell line. Expression of each dCas9 construct was introduced by transduction with a specific Cas9-expressing lentivirus: pG1H125_dCas9-Blast (Addgene #85417) for dCas9; pHR-SFFV-KRAB-dCas9-P2A-mCherry (Addgene #60954) for dCas9-KRAB; Lenti-hEF1-BLAST-dCas9-VPR (Dharmacon, CAS11916) for dCas9-VPR. Successfully transduced cells were then selected for by mCherry expression (dCas9-KRAB) or treatment with 10 μg/mL blasticidin (dCas9 and dCas9-VPR; Gibco). Cells were then seeded into 24-well plates at a density of 50,000 cells per well. 100 μL of sgRNA lentivirus was added. After 24 h, media was replaced and after 48 h cells were lysed using the Cells-to-Ct kit (Life Technologies) for subsequent gene expression analysis by RT-PCR.

Real-time PCR

Real-time PCR analysis of gene expression in cDNA samples was performed using Taqman probes (Life Technologies) for IGFBP2 (MIM: 146731), IGFBPS, and RPL37A (MIM: 613314) normalized to the housekeeping gene GAPDH (ThermoFisher; Hs01040719_m1, IGFBPS: Hs00181213_m1, RPL37A: Hs01102345_m1, GAPDH: Hs03929097_g1). Reactions of 5 μL were performed using Taqman Universal Master Mix II, without UNG (Applied Biosystems) according to the manufacturer’s instructions.

Statistical analysis of reporter gene assays and CRISPR-based enhancer perturbation

Firefly luciferase activity was internally normalized to renilla luciferase activity, and each condition normalized to the “IGFBP promoter-alone” (IGFBP5-PROM) construct. Setting IGFBP5-PROM to 1.0, for each putative enhancer-containing reporter gene construct we used t tests to test (1) H0: the mean dual luciferase ratio does not differ from 1.0 and (2) H0: the ALT construct does not differ from the REF construct. To compare mean dual luciferase ratios for each combination of SNP and SV at PRE2, we used three-way analysis of variance adjusting each mean dual luciferase ratios for each combination of SNP and SV variant for all other variants. To account for multiple testing, we used a Bonferroni corrected p value of 0.0056 (individual comparisons, Table S4). Fine-scale mapping of this region in women of Asian Ancestry (12,481 affected individuals and 12,758 control subjects) did not identify any population-specific signals (all associations p > 5 × 10⁻⁸; Figure S3). None of the credible causal variants at signal 2 was present in women of Asian ancestry. The published causal variant at signal 1 (rs4442975) and all of the signal 3 credible causal variants (Table S5) were nominally associated with breast cancer risk in Asian women (p < 0.05). At signal 3, the index variants differ between Europeans and Asians (rs5838651 and rs138522813; 21,826,5367, respectively) but none of the European credible causal variants could be excluded on the basis of the Asian data.

The T-allele of rs4442975 was associated with reduced breast cancer risk (carrier OR = 0.88, 95% CI 0.87–0.89, p = 1.3 × 10⁻⁷⁵ and OR = 0.94, 95% CI 0.89–1.00, p = 0.04 in European and Asian women, respectively) and the delG-allele of rs5838651 was associated with increased risk (carrier OR = 1.07, 95% CI 1.05–1.08, p = 1.5 × 10⁻¹⁶ and OR = 1.07, 95% CI 1.03–1.11, p = 0.0008 in European and Asian women, respectively; Table 1). The delT-allele of rs138522813 was associated with reduced risk (carrier OR = 0.80 95% CI 0.77–0.83, p = 5.5 × 10⁻³⁵). Stratifying by ER status, the signal 1 (rs4442975) and signal 2 (rs138522813) SNPs were more strongly associated with ER⁺ disease; for the signal 3 SNP (rs5838651), there was no evidence that the ORs differed by ER status (Table S6).

Prioritization of credible variants for functional follow up

Fachal and colleagues3 used a Bayesian approach (PAIN-TOR) that combines genetic association, linkage disequilibrium, and enriched genomic features to determine variants with high posterior probabilities of being causal (Table S4).20 rs4442975, the only credible causal variant at signal 1 (posterior probability = 0.84), has previously been proposed to have a functional effect on breast cancer risk.11,12 Four of the five variants at signal 2 had posterior probabilities ≥ 0.20 (combined posterior probability 0.997); none of the variants at signal 3 had posterior probabilities > 0.15. To further prioritize putative causal variants at signals 2 and 3, we aligned the 47 credible variants at these signals with markers of open chromatin (DNase I), active transcription (F300), active
enhancers (H3K27Ac, H3K4me1), and breast-relevant TFs (FOXA1, GATA3, ERα) generated in T-47D and MCF-7 breast cancer cells \textsuperscript{15–17} (Table S4). Consistent with the PAINTOR posterior probabilities, four variants at signal 2 colocalized with at least one of these features. In addition, we identified two variants at signal 3 that colocalized with one of these features. These six variants were prioritized for further functional annotation.

**Reporter gene assays of prioritized variants**

For SNPs, we generated reference (REF) and alternative (ALT) constructs in which the putative regulatory element, defined in the first instance as a 500 to 700 bp region centered on the SNP or SNP pair (PRE2A rs572022984; PRE2B rs199804270 and rs72951831; PRE3 rs12694417 and rs12988242), was cloned upstream of a luciferase reporter gene, driven by the IGFBP5 promoter (Figure S1). For the structural variant esv3594306, which is defined by the presence (REF) or absence (ALT) of a 1.4 kb region (chr2:217,955,891–217,957,273; GRCh37/hg19), we generated separate REF constructs for PRE2A and PRE2B and a single ALT construct in which the centromeric sequences at PRE2A were juxtaposed to the telomeric sequences at PRE2B with the intervening 1.4 kb deleted (Figure 1B). Comparing the REF construct at each region with the IGFBP5 promoter construct (IGFBP5-PROM), there was evidence that two of the putative regulatory elements (PRE2B and PRE3) enhanced transcription from the IGFBP5 promoter (Figure 2). For PRE2B, both alleles demonstrated...
strong enhancer activity (PRE2B-REF/REF: fold change \[FC\] = 27.9, \(p = 0.004\); PRE2-DEL-ALT/ALT: FC = 50.5, \(p = 0.004\)). For PRE3 the activity was more modest and only significant (\(p < 0.0056\); Material and methods) for the ALT allele in T-47D (PRE3-REF/REF: FC = 1.8, \(p = 0.03\); PRE3-ALT/ALT FC = 2.2, \(p = 0.008\); Figure 2). To test these constructs for cell type specificity, we used HepG2 (hepatocyte carcinoma), 293T (embryonic kidney), and HCT116 (colorectal carcinoma) cells. The only construct that influenced transcription from the \(IGFBP5\) promoter in these non-breast cells was PRE2DEL-ALT/ALT in 293T cells and with an effect size that was an order of magnitude lower (FC = 1.9, \(p = 0.002\); Figure S4) compared to the breast cancer cell lines (FC > 40; Figure 2). Comparing ALT constructs with REF constructs, only the PRE2 region showed a significant difference between alleles, with the (protective) PRE2-DEL-ALT/ALT allele being associated with greater activity than PRE2B-REF/REF allele (MCF-7 \(FC = 1.8\), \(p = 0.003\); T-47D \(FC = 1.6\), \(p = 0.09\); Figure 2).

| rs4442975 | rs138522813 | rs5838651 |
|-----------|-------------|----------|
| Europeans |            |          |
| MAF       | Cases       | Controls  |
| rs4442975 | 0.49        | 36,471   |
| rs138522813 | 0.035      | -        |
| rs5838651 | 0.61        | -        |

| Asians    |            |          |
| MAF       | Cases       | Controls  |
| rs4442975 | 0.87        | 4,994    |
| rs138522813 | -          | -        |
| rs5838651 | 0.61        | -        |

*MAF, minor allele frequency

\(OR\), per allele odds ratio

\(P_1\), test of \(H_0\) no association between SNP and breast cancer risk

\(P_{\text{het1}}\), test of \(H_0\) no difference between iCOGS and OncoArray data

\(P_{\text{het2}}\), test of \(H_0\) no difference between European and Asian data

\(rs138522813\)-Del allele is extremely rare in Asians (MAF ~ 0.05%) and not analyzed in Asian data

The PRE containing the reference (REF) allele at each SNP was cloned downstream of the \(IGFBP5\) promoter to generate reference (REF) luciferase constructs. Alternative (ALT) alleles were generated by site-directed mutagenesis. Coordinates of the PREs are given in Table S2, diagrams are in Figure S1. Error bars denote standard deviations based on three independent experiments each done in triplicate. \(p\) values were determined by t tests and a Bonferroni correction was applied to account for multiple testing.
PRE2B-REF/REF allele, this time in T-47D cells (MCF-7 FC = 1.5, p = 0.15; T-47D FC = 2.7, p = 0.002; Figure S5A).

The PRE2DEL-ALT/ALT construct comprises a haplotype of three tightly linked variants: the ALT alleles of the two SNPs (rs199804270:GA:G, rs72951831:G:T) with the ALT (deletion) allele of the structural variant (esv3594306) that brings two separate ERα, FOXA1, GATA3, and P300 ChIP-seq peaks into juxtaposition (Figure 1B). To differentiate individual effects, each allele of each SNP was introduced onto esv3594306 insertion and deletion backgrounds separately using site-directed mutagenesis. The PRE2A SNP (rs572022984) was not considered further due to technical issues (Material and methods). In a combined analysis, adjusting each variant for the other two variants, there was evidence that deletion constructs consistently showed greater activity than insertion constructs (MCF-7: DEL FC = 43.4, INS FC = 34.4, i.e., average additional FC for DEL = 1.3, \(p_{\text{het}} = 0.01\); T-47D: DEL FC = 47.3, INS FC = 21.6, i.e., average additional FC for DEL = 2.2, \(p_{\text{het}} = 1.7 \times 10^{-6}\); Figure 3).

**Figure 3. Luciferase reporter assays following transient transfection of constructs with allelic variants at PRE2B and PRE2DEL into MCF-7 and T-47D cells**

Reporter gene constructs with all possible combinations of rs199804270 and rs72951831 and esv3594306 were generated by site-directed mutagenesis of the naturally occurring haplotypes at PRE2B and PRE2DEL (Material and methods) into MCF-7 (A) and T-47D (B) cells. Coordinates of the PREs are given in Table S2, diagrams are in Figure S1. Error bars denote standard deviations based on three independent experiments each done in triplicate. 3-way ANOVA was used to compare each variant, adjusted for the other two variants, a Bonferroni correction was applied to account for multiple testing. *p < 0.017, **p \leq 0.0017.

**CRISPR-based perturbation of PRE2**

Reporters gene assays do not reflect the “normal” genomic context of a regulatory element. Specifically, the assay tests whether the putative regulatory element can influence expression in an episomal context and from a distance of a few kilobases; in vivo, PRE2 maps approximately 400 kb from the IGFBP5 promoter, colocalizing with the transcription start site and likely reduces expression of IGFBP5 by steric hindrance (60% reduction, \(p = 0.004\); Figure S7A). In the CRISPRi setting, all three sgRNAs targeting the IGFBP5 promoter repressed IGFBP5 expression significantly to 8%-15% of levels in the negative controls (\(p = 0.001\), \(p = 0.001\), and \(p = 0.0008\) for guides PROM-1 to -3, respectively; Figure S8A). No sgRNA targeting non-promoter sequences influenced IGFBP5 expression (Figures S8A and S8B). In the CRISPRa setting, the sgRNA 5′ to the IGFBP5 promoter (PROM-3; Figure 4A) enhanced IGFBP5 expression more than 60-fold (\(p = 0.0008\)) and the PRE-1-positive control sgRNA (PRE1-1) targeting rs4442975 also enhanced IGFBP5 expression (FC = 3.7, \(p = 0.006\); Figure 4A). In addition, four of the nine sgRNAs targeting sequences at PRE2 enhanced IGFBP5 expression; specifically PRE2-1 and -2 targeting the ERα ChIP-seq peak at the centromeric deletion breakpoint (PRE2-1: FC = 3.7, \(p = 0.0005\); PRE2-2: FC = 3.1, \(p = 0.001\)), PRE2-5 at the distal end of the deletion region (PRE2-5: FC = 3.2, \(p = 0.002\)), and PRE2-8 targeting the ERα ChIP-seq peak immediately telomeric to the deletion region (PRE2-8: FC = 5.3, \(p = 0.002\); Figures 4B and S5A). None of the sgRNAs influenced expression of two genes mapping immediately 3′ to IGFBP5 (IGFBP2 and RPL37A; Figure 4C).
Discussion

Fine-scale mapping at the 2q35 breast cancer locus in women of European ancestry\(^3\) confirmed rs4442975 as the probable causal variant at signal 1 and reduced the number of credible causal variants at signal 2 from 14 to 5;\(^{3,14}\) at signal 3, however, there remained 42 credible causal variants that could not be excluded as causal on statistical grounds alone in either the European or the Asian data. Low-throughput functional approaches that are used to investigate putative causal variants, including reporter gene assays and CRISPR screens, become prohibitive with large numbers of credible causal variants and most single locus\(^{11,14,24–38}\) and global\(^{1,6}\) annotation studies have used co-localization of credible causal variants with markers of open chromatin, active histone modifications, and transcription factor binding in relevant cell types to prioritize credible causal variants for functional follow up. Of the 811 annotation tracks that were examined in a recent global fine-scale mapping analysis,\(^3\) credible causal variants were enriched at three types of genomic features that are relevant to long-range regulatory elements: (1) open chromatin in ER\(^+\) cell lines and normal breast, (2) the active histone marks H3K4me1 and H3K27ac in MCF-7 cells, and (3) ESR1, FOXA1, GATA3, and P300 TF binding sites. By aligning the five credible causal variants at PRE2 and the 42 credible causal variants at PRE3 with these marks (Table S4), we were able to prioritize 4 of the 5 credible causal variants at PRE2 and 2 of the 42 credible causal variants at PRE3 for follow-up studies. By taking this approach there is, inevitably, the possibility that we have excluded one or more causal variants from our follow-up analyses. For PRE2 this seems unlikely as we selected four out of the five credible causal variants for further follow-up studies. For PRE3 it is entirely possible, or even probable, that we failed to prioritize one or more causal variant(s); improving our ability to discriminate more accurately between potentially functional variants and large numbers of correlated variants will require genome-wide datasets with functional outputs\(^{21,39,40}\) generated in more relevant cellular disease models and taking advantage of single-cell technologies.\(^1\)

Using reporter gene assays, we have demonstrated that both the distal region of PRE2 (PRE2B) and the entire PRE3 region can enhance transcription from the IGFBP5 promoter in a cell-type-specific manner. Despite co-localizing with multiple markers, we found no evidence that...
the proximal region of PRE2 (PRE2A) acts as an independent enhancer element. The ChIP-seq peaks at this region are, however, relatively weak (Figure 1B); combining data from both PRE2A alleles, in both breast cancer cell lines to increase our power (i.e., using 12 replicates rather than 3) the overall mean fold change for PRE2A was 1.14 (1.03–1.26, \( p = 0.01 \)), consistent with the presence of a very modest enhancer element. Comparing REF constructs with ALT constructs, we found no evidence that either of the credible causal variants at PRE3 (rs12694417, rs12988242) altered the activity of the PRE. This does not exclude these SNPs as functional; as above, modest effects on enhancer activity may be difficult to detect and variants that, for example, influence chromatin accessibility may not be detectable in transient assays.11 However, without preliminary in vitro evidence to suggest that one of these variants alters cell-type-specific transcription from the IGFBP5 promoter, pursuing further functional studies that are predicated on this very assumption seems unlikely to be fruitful. By contrast, one comparison that was consistent and significant between constructs and across the two breast cancer cell lines was that PRE2 deletion alleles had stronger enhancer activity than PRE2 insertion alleles.

The purpose of our CRISPR-based enhancer perturbation was 2-fold: specifically, to interrogate the PRE2 region within its normal genomic context and more generally to evaluate CRISPRi and CRISPRa approaches for interrogating long-range regulatory elements that harbor credible causal variants. As none of our PRE2 sgRNAs impacted IGFBP5 expression significantly in the CRISPRi setting, our analysis raises questions as to the utility of this approach for characterizing long-range regulatory elements (PRE2 maps approximately 400 kb telomeric to the IGFBP5 promoter). This is at odds with results of a systematic CRISPRi screen to identify enhancer elements in K562 cells, which demonstrated CRISPRi-mediated repression of c-MYC expression by sgRNAs targeting sequences mapping up to 1.9 Mb downstream of c-MYC.22 In this analysis, however, CRISPRi-mediated repression by these distal elements was modest compared to CRISPRi-mediated repression by more proximal elements and, even based on 12 biological replicates, of borderline statistical significance.22 By contrast, using CRISPRa we were able to confirm that one or more elements within PRE2 can act as a long-range regulatory element that specifically targets IGFBP5 (rather than IGFBP2 or RPL37A). Four of the nine guide RNAs targeting dCas9-VPR to sequences at PRE2 increased expression of IGFBP5; three of these colocalized with ER\( \alpha \), FOXA1, and GATA3 ChIP-seq peaks (PRE2-1, -2,
and -8) and a fourth (PRE2-5) mapped within the esv3594306 deleted region (Figure 5A). There were also two guides which targeted dCas9-VPR to sequences that map close to the distal ER, FOXA1, and GATA3 ChIP-seq peak (PRE2-6 and -7) but did not increase IGFBP5 expression; this may reflect the very variable efficiency of different guide RNAs. We present a theoretical model in which we hypothesize that all of the PRE2 guides that increased expression of IGFBP5 increased the local density of activating TF domains by bringing a VPR domain into the proximity of a cluster of TF ChIP-seq peaks; one implication of the increase in IGFBP5 expression we observed with PRE2-5, which maps approximately 450 bp from the center of the nearest cluster of ChIP-seq peaks (Figure 5A), is that these regulatory elements may extend over relatively large (>1 kb) regions. This should not, perhaps, be surprising; at a subset of strongly activated E2-responsive enhancers, it has previously been shown that ERα recruits DNA-binding transcription factors in trans, to form a large (1–2 MDa) complex. 

It has previously been suggested that sequences mapping to PRE2 act as a repressor element which, in the presence of low-dose estradiol, acts to reduce IGFBP5 expression. By contrast, our data support PRE2 acting as a powerful enhancer element with the deletion allele increasing expression of IGFBP5 over and above that of the insertion allele with or without estradiol stimulation. Overall, our data are consistent with a hypothetical model in which the juxtaposition of the two ERα, FOXA1, GATA3 binding sites at PRE2 by deletion of approximately 1.4 kb of intervening sequence generates a single extended binding region (Figure 5B) that is causally associated with increased enhancer activity, higher levels of expression of the putative tumor suppressor gene IGFBP5, and a reduction in breast cancer risk (OR = 0.77, p = 2.2 × 10^{-29}) that is largely restricted to ER+ disease.

In conclusion, we have identified putative enhancer elements at two additional 2q35 breast cancer risk loci. One of these, mapping approximately 400 kb telomeric to IGFBP5, enhances transcription from the IGFBP5 promoter by a factor of 30- to 40-fold. For this element we provide evidence that a deletion of 1.4 kb is causally associated with increased enhancer activity and suggest a mechanism for this increased activity.

Data and code availability
Summary results for all variants genotyped by the Breast Cancer Association Consortium BCAC (including rs45446698) are available at http://bcac.ccge.medschl.cam.ac.uk/. Requests for data can be made to the corresponding author or the Data Access Coordination Committee (DACC) of the Breast Cancer Association Consortium via email to BCAC@medschl.cam.ac.uk.

Supplemental information
Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.05.013.

Consortia
The NBCS Collaborators are Anne-Lise Børresen-Dale, Grethe I. Grenaker Alnaes, Kristine K. Sahlberg, Lars Ottestad, Rolf Kåresen, Ellen Schlichting, Marit Muri Holmen, Toril Sauer, Vilde Haakensen, Olav Engebråten, Bjørn Naume, Alexander Fosså, Cecile E. Kiørured, Kristin V. Reinersen, Åslaug Helland, Margit Riis, Jürgen Geisler, and OSBREAC. The kConFab Investigators are David Amor, Lesley Andrews, Yoland Antill, Rosemary Baleline, Jonathan Beesley, Ian Bennett, Michael Bogwitz, Leon Botes, Meagan Brennan, Melissa Brown, Michael Buckley, Jo Burke, Phyllis Butow, Liz Caldnon, Ian Campbell, Deepa Chauhan, Manisha Chauhan, Georgia Chenevix-Trench, Alice Christopher, Paul Cohen, Alison Colley, Ashley Crook, James Cui, Margaret Cummings, Sarah Jane Dawson, Anna defazio, Martin Delatycki, Rebecca Dickson, Joanne Dixon, Ted Edkins, Stacey Edwards, Gelareh Farshid, Andrew Fellows, Georgina Fenton, Michael Field, James Flanagan, Peter Fong, Laura Forrest, Stephen Fox, Juliet French, Michael Friedlander, Clara Gaff, Mike Gattas, Peter George, Sian Greening, Marion Harris, Stewart Hart, Nick Hayward, John Hopper, Cass Hoskins, Clare Hunt, Paul James, Mark Jenkins, Alexa Kidd, Judy Kirk, Jessica Koehler, James Kollias, Sunil Lakhan, Mitchell Lawrence, Geoff Lindeman, Lara Lipton, Liz Lobb, Graham Mann, Deborah Marsh, Sue Anne Mcclachlan, Bettina Meiser, Roger Milne, Sophie Nightingale, Shona O’Connell, Sarah O’Sullivan, David Gallego Ortega, Nick Pachter, Briony Patterson, Amy Pearn, Kelly Philips, Ellen Pieper, Edwina Rickard, Bridget Robinson, Mona Saleh, Elizabeth Salisbury, Christobel Saunders, Jodi Saunus, Rodney Scott, Clare Scott, Adrienne Sexton, Andrew Shelling, Peter Simpson, Melissa Southey, Amanda Spurdle, Jessica Taylor, Renea Taylor, Heather Thorne, Alison Trainer, Kathy Tucker, Jane Visvader, Logan Walker, Rachael Williams, Ingrid Winship, and Mary Ann Young. The ABCTB Investigators are Christine Clarke, Deborah Marsh, Rodney Scott, Robert Baxter, Desmond Yip, Jane Carpenter, Alison Davis, Nirama Pathmanathan, Peter Simpson, Dinny Graham, and Mythily Sachchithananthan.

Declaration of interests
M.W.B. conducts research funded by Amgen, Novartis, and Pfizer. P.A.F. conducts research funded by Amgen, Novartis, and Pfizer and received honoraria from Roche, Novartis, and Pfizer. A.W.K. received research funding to her institution from Myriad Genetics for an unrelated project (funding dates 2017–2019). U.M. has stockownership in Abcodia Ltd. All other authors declare no conflict of interest.

Received: October 20, 2020
Accepted: May 25, 2021
Published: June 18, 2021

Web resources
1000 Genomes Project, https://www.internationalgenome.org/

References
1. Lichou, F., and Trynka, G. (2020). Functional studies of GWAS variants are gaining momentum. Nat. Commun. 11, 6283.
2. Monteiro, A.N., and Freedman, M.L. (2013). Lessons from postgenome-wide association studies: functional analysis.
of cancer predisposition loci. J. Intern. Med. 274, 414–424.

3. Fachal, L., Aschard, H., Beesley, J., Barnes, D.R., Allen, J., Kar, S., Pooley, K.A., Dennis, J., Michaillidou, K., Turman, C., et al.; GEMO Study Collaborators; EMBRACE Collaborators; kConFab Investigators; HEBON Investigators; and ABCTB Investigators (2020). Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. Nat. Genet. 52, 56–73.

4. Consortium, E.P.; and ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.

5. Michaillidou, 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; and 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. Michaillidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., et al.; NBCS Collaborators; ABCTB Investigators; and ConFab/AOCS Investigators (2017). Association analysis identifies 65 new breast cancer risk loci. Nature 551, 92–94.

7. Milne, R.L., Kuchenbaecker, K.B., Michaillidou, K., Beesley, J., Kar, S., Lindström, S., Hui, S., Lemaçon, A., Soucy, P., Dennis, J., et al.; ABCTB Investigators; EMBRACE; GEMO Study Collaborators; HEBON; kConFab/OAOC Investigators; and NBCS Collaborators (2017). Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. Nat. Genet. 49, 1767–1778.

8. Zhang, H., Ahearn, T.U., Lecarpentier, J., Barnes, D., Beesley, J., Qi, G., Jiang, X., O’Mara, T.A., Zhao, N., Bolla, M.K., et al.; kConFab Investigators; ABCTB Investigators; EMBRACE Study; and GEMO Study Collaborators (2020). Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat. Genet. 52, 572–581.

9. Garcia-Closas, M., Couch, F.J., Lindstrom, S., Michaillidou, K., Schmidt, M.K., Broek, M.N., Orr, N., Rhode, S.K., Riboli, E., Feigelson, H.S., et al.; Gene ENvironmental Interaction and breast Cancer (GENICA) Network; kConFab Investigators; Familial Breast Cancer Study (FBCS); and Australian Breast Cancer Tissue Bank (ABCTB) Investigators (2013). Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat. Genet. 45, 392–398, e1–e2.

10. Stacey, S.N., Manolescu, A., Sulem, P., Rafnson, T., Gudmundsson, J., Gudjonssohn, S.A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason, A., et al. (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat. Genet. 39, 865–869.

11. Ghoussaini, M., Edwards, S.L., Michaillidou, 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; and Australian Ovarian Cancer Management Group (2014). Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. Nat. Commun. 4, 4999.

12. Dryden, N.H., Broome, I.R., Dudbridge, F., Johnson, N., Orr, N., Schoenfelder, S., Nagano, T., Andrews, S., Wingett, S., Kozarewa, I., et al. (2014). Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. Genome Res. 24, 1854–1868.

13. Baxter, J.S., Leavy, O.C., Dryden, N.H., Maguire, S., Johnson, N., Fedele, V., Simigdal, N., Martin, L.A., Andrews, S., Wingett, S.W., et al. (2018). Capture Hi-C identifies putative target genes at 33 breast cancer risk loci. Nat. Commun. 9, 1028.

14. Wysszynski, A., Hong, C.C., Lam, K., Michaillidou, K., Lytle, C., Yao, S., Zhang, Y., Bolla, M.K., Wang, Q., Dennis, J., et al.; GENICA Network; kConFab Investigators; and Australian Ovarian Cancer Study Group (2016). An intergenic risk locus containing an enhancer deletion in 2q35 modulates breast cancer risk by deregulating IGFBP5 expression. Hum. Mol. Genet. 25, 3863–3876.

15. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature 489, 75–82.

16. Gertz, J., Savic, D., Varley, K.E., Partridge, E.C., Safi, A., Jain, P., Cooper, G.M., Reddy, T.E., Crawford, G.F., and Myers, R.M. (2013). Distinct properties of cell-type-specific and shared transcription factor binding sites. Mol. Cell 52, 25–36.

17. Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature 498, 516–520.

18. Swerdlow, A.J., Jones, M.E., Schoemaker, M.J., Hemmings, J., Thomas, D., Williamson, J., and Ashworth, A. (2011). The Breakthrough Generations Study: design of a long-term UK cohort study to investigate breast cancer etiology. Br. J. Cancer 105, 911–917.

19. Ran, E.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308.

20. Kichaev, G., Yang, W.Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A.L., Kraft, P., and Pasaniuc, B. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. 10, e1004722.

21. Gordon, M.G., Inoue, F., Martin, B., Schubach, M., Agarwal, V., Whalen, S., Feng, S., Zhao, J., Ashuach, T., Ziffra, R., et al. (2020). IntentiMPRA and MPRAFlow for high-throughput functional characterization of gene regulatory elements. Nat. Protoc. 15, 2387–2412.

22. Fulco, C.P., Munschauer, M., AnyoHa, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science 354, 769–773.

23. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154, 442–451.

24. Guo, X., Long, J., Zeng, C., Michaillidou, K., Ghoussaini, M., Bolla, M.K., Wang, Q., Milne, R.L., Shu, X.O., Cai, Q., et al.; kConFab Investigators (2015). Fine-scale mapping of the 4q24 locus identifies two independent loci associated with breast cancer risk. Cancer Epidemiol. Biomarkers Prev. 24, 1680–1691.

25. Glubb, D.M., Maranian, M.J., Michaillidou, K., Pooley, K.A., Meyer, K.B., Kar, S., Carlebur, S., O’Reilly, M., Betts, J.A., Hillman, K.M., et al.; GENICA Network; kConFab Investigators;
and Norwegian Breast Cancer Study (2015). Fine-scale mapping of the 9q11.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. Am. J. Hum. Genet. 96, 5–20.

26. Dunning, A.M., Michailidou, K., Kuchenbaecker, K.B., Thompson, D., French, J.D., Beesley, J., Healey, C.S., Kar, S., Pooley, K.A., Lopez-Knowles, E., et al.; EMBRACE; GEMO Study Collaborators; HEBON; and kConFab Investigators (2016). Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. Nat. Genet. 48, 374–386.

27. Shi, J., Zhang, Y., Zheng, W., Michailidou, K., Ghoussaini, M., Bolla, M.K., Wang, Q., Dennis, J., Lush, M., Milne, R.L., et al.; Mervi Grip; and kConFab Investigators (2016). Fine-scale mapping of 8q24 locus identifies multiple independent risk variants for breast cancer. Int. J. Cancer 139, 1303–1317.

28. Orr, N., Dudbridge, F., Dryden, N., Maguire, S., Novo, D., Perrakis, E., Johnson, N., Ghoussaini, M., Hopper, J.L., Southey, M.C., et al.; GENICA Network; kConFab Investigators; and Australian Ovarian Cancer Study Group (2015). Fine-mapping identifies two additional breast cancer susceptibility loci at 9q31.2. Hum. Mol. Genet. 24, 2966–2984.

29. Darabi, H., McCue, K., Beesley, J., Michailidou, K., Nord, S., Kar, S., Humphreys, K., Thompson, D., Ghoussaini, M., Bolla, M.K., et al.; German Consortium of Hereditary Breast and Ovarian Cancer; and kConFab Investigators (2015). Polymorphisms in a Putative Enhancer at the 10q21.2 Breast Cancer Risk Locus Regulate NRBF2 Expression. Am. J. Hum. Genet. 97, 22–34.

30. Meyer, K.B., O’Reilly, M., Michailidou, K., Carlebur, S., Edwards, S.L., French, J.D., Prathallingham, R., Dennis, J., Bolla, M.K., Wang, Q., et al.; GENICA Network; kConFab Investigators; and Australian Ovarian Cancer Study Group (2013). Fine-scale mapping of the FGF22 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. Am. J. Hum. Genet. 93, 1046–1060.

31. Lightfoot, S.J., Moradi Marjaneh, M., Al-Ejeh, F., Lim, Y.C., Shi, W., Sivakumaran, H., Tropée, R., Patch, A.M., Clark, M.B., Bartonicek, N., et al. (2017). Long Noncoding RNAs CUPID1 and CUPID2 Mediate Breast Cancer Risk at 11q13 by Modulating the PRC2 Phenotype Complex. Cell 171, 522–538.

32. French, J.D., Ghoussaini, M., Edwards, S.L., Meyer, K.B., Michailidou, K., Ahmed, S., Khan, M., Maranian, M.J., O’Reilly, M., Hillman, K.M., et al.; GENICA Network; and 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.

33. Ghoussaini, M., French, J.D., Michailidou, K., Nord, S., Beesley, J., Canisus, S., Hillman, K.M., Kaufmann, S., Sivakumaran, H., Moradi Marjaneh, M., et al.; kConFab/AOCS Investigators; and NBCS Collaborators (2016). Evidence that the Sp12 Variant rs10941679 Confers Susceptibility to Estrogen-Receptor-Positive Breast Cancer through FGFR10 and MRPS30 Regulation. Am. J. Hum. Genet. 99, 903–911.

34. Horne, H.N., Chung, C.C., Zhang, H., Yu, K., Prokunina-Olsson, L., Michailidou, K., Bolla, M.K., Wang, Q., Dennis, J., Hopper, J.L., et al.; kConFab/AOCS Investigators (2016). Fine-Mapping of the 1p11.2 Breast Cancer Susceptibility Locus. PLoS ONE 11, e0160316.

35. Zeng, C., Guo, X., Long, J., Kuchenbaecker, K.B., Droit, A., Michailidou, K., Ghoussaini, M., Kar, S., Freeman, A., Hopper, J.L., et al.; EMBRACE; behalf of GEMO Study Collaborators; HEBON; KConFab; and AOCS Investigators (2016). Identification of independent association signals and putative functional variants for breast cancer risk through fine-scale mapping of the 12p11 locus. Breast Cancer Res. 18, 64.

36. Lin, W.Y., Camp, N.J., Ghoussaini, M., Beesley, J., Michailidou, K., Hopper, J.L., Apicella, C., Southey, M.C., Stone, J., Schmidt, M.K., et al.; GENICA Network; kConFab Investigators; Australian Ovarian Cancer Study Group; and Breast and Ovarian Cancer Susceptibility (BOCS) Study (2015). Identification and characterization of novel associations in the CASP8/AL52CR12 region on chromosome 2 with breast cancer risk. Hum. Mol. Genet. 24, 285–298.

37. Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K., Tyer, J.P., Edwards, S.L., Pickett, H.A., Shen, H.C., Smart, C.E., et al.; Australian Cancer Study; Australian Ovarian Cancer Study; Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (kConFab); Gene Environment Interaction and Breast Cancer (GENICA); Swedish Breast Cancer Study (SWE-BRCA); Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON); Epidemiological study of BRCA1 & BRCA2 Mutation Carriers (EMBRACE); and Genetic Modifiers of Cancer Risk in BRCA1/2 Mutation Carriers (GEMO) (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nat. Genet. 45, 371–384, e1–e2.

38. Lawrenson, K., Kar, S., McCue, K., Kuchenbaecker, K., Michailidou, K., Tyer, J., Beesley, J., Ramus, S.J., Li, Q., Delgado, M.K., et al.; GEMO Study Collaborators; EMBRACE; Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON); KConFab Investigators; and Australian Ovarian Cancer Study Group (2016). Functional mechanisms underlying pleiotropic risk alleles at the 19p13.1 breast-ovarian cancer susceptibility locus. Nat. Commun. 7, 12675.

39. Inoue, F., and Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. Genomics 106, 159–164.

40. Arnold, C.D., Gerlach, D., Stelzer, C., Borys, A.C., Goding, C.R., and Vance, K.W. (2019). The MITF-SOX10 regulated long non-coding RNA DIRC3 is a melanoma tumour suppressor. PLoS Genet. 15, e1008501.