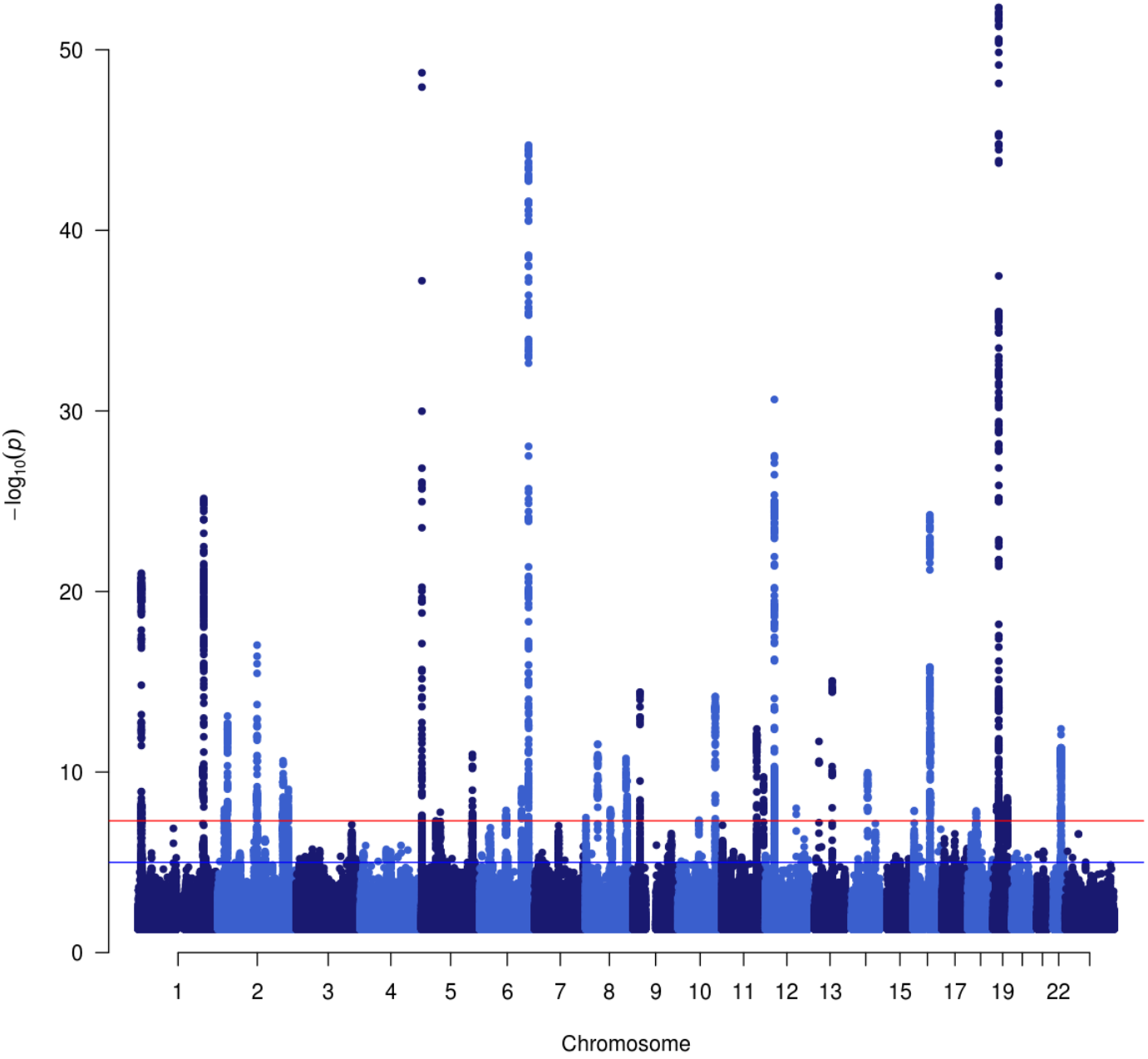
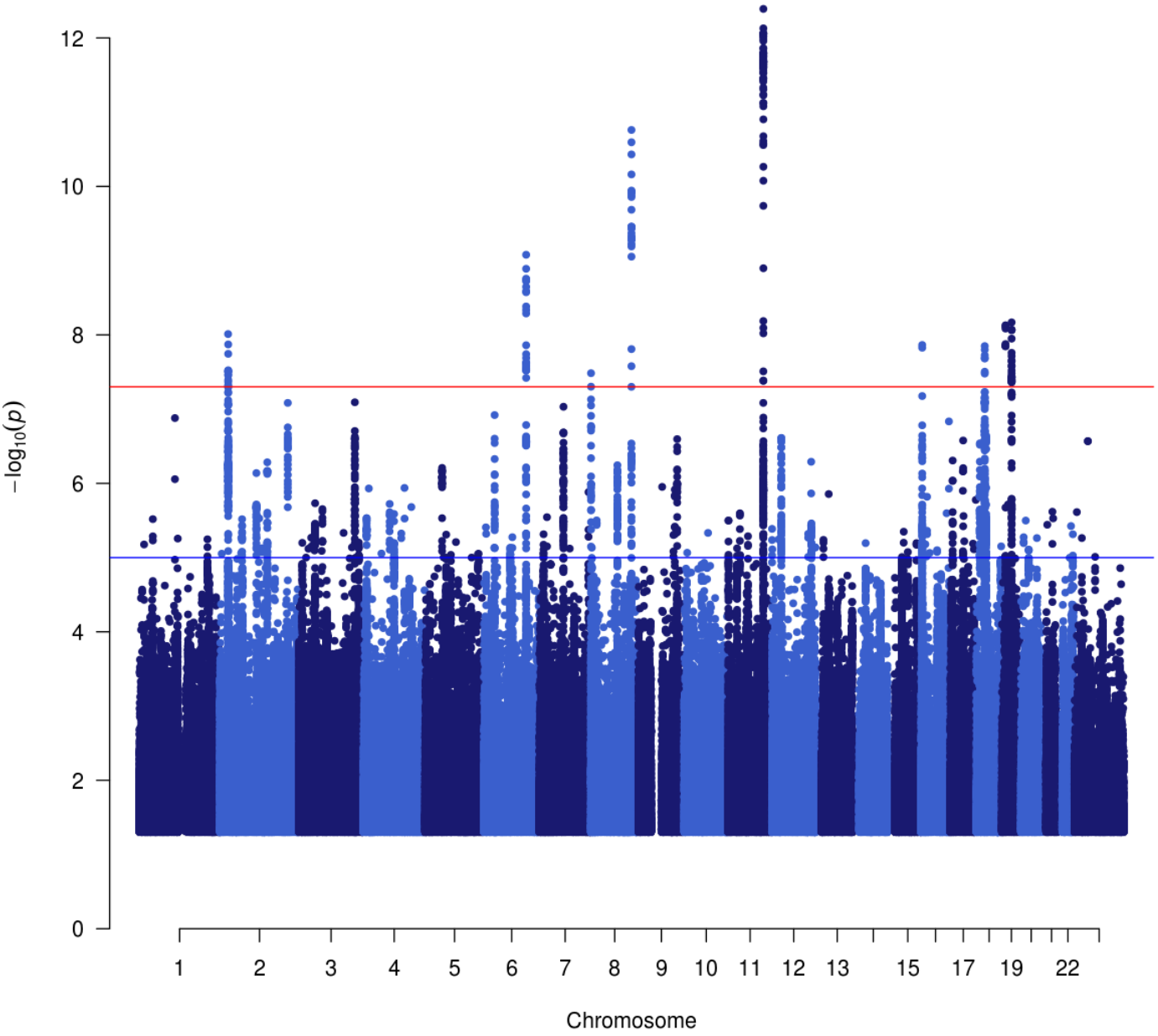


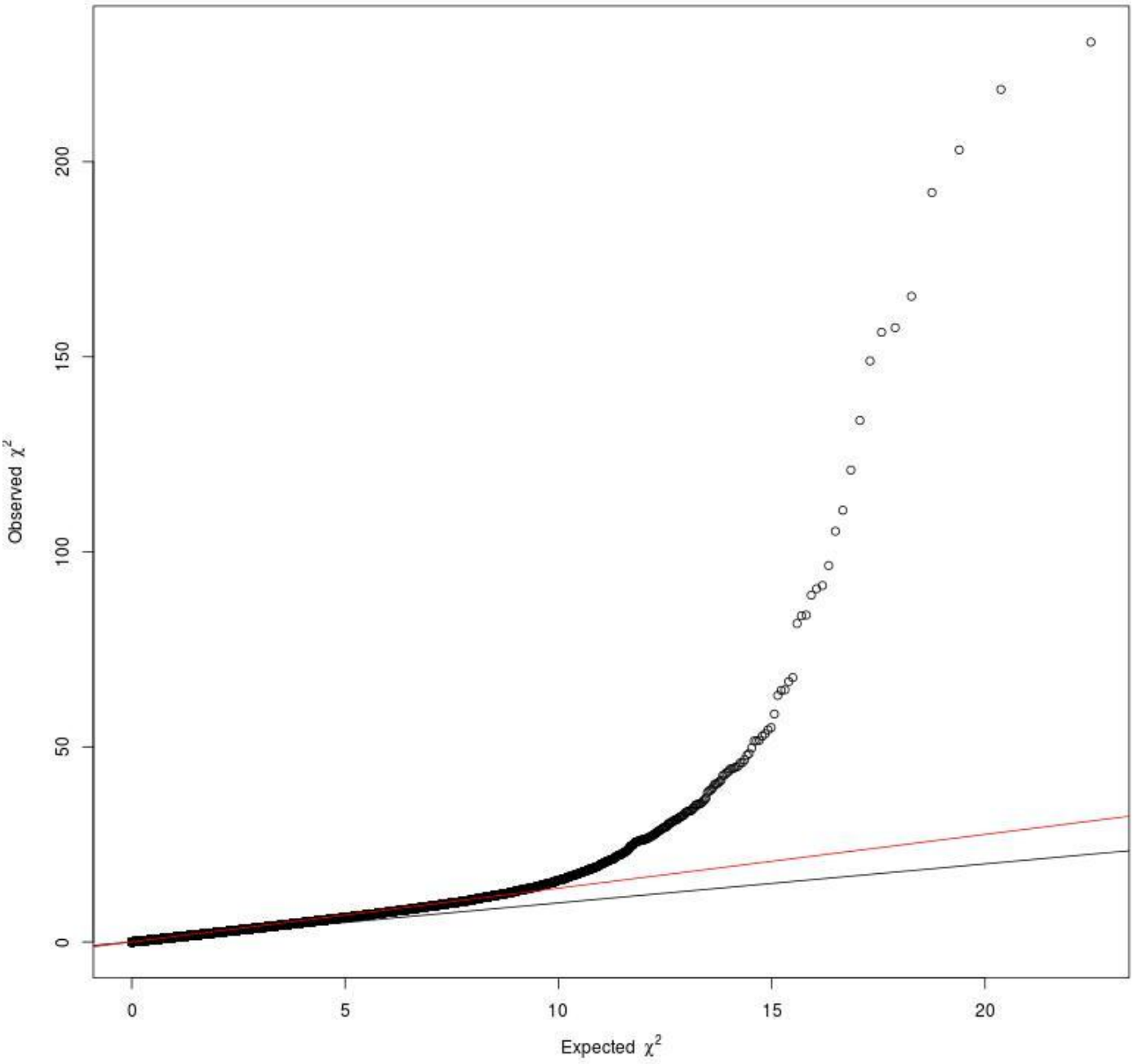
Supplementary Figure 1

Manhattan plot of associations with breast cancer risk for all imputed and genotyped SNPs using combined data from ER-negative cases and controls and *BRCA1* mutation carriers, before excluding known breast cancer susceptibility loci.



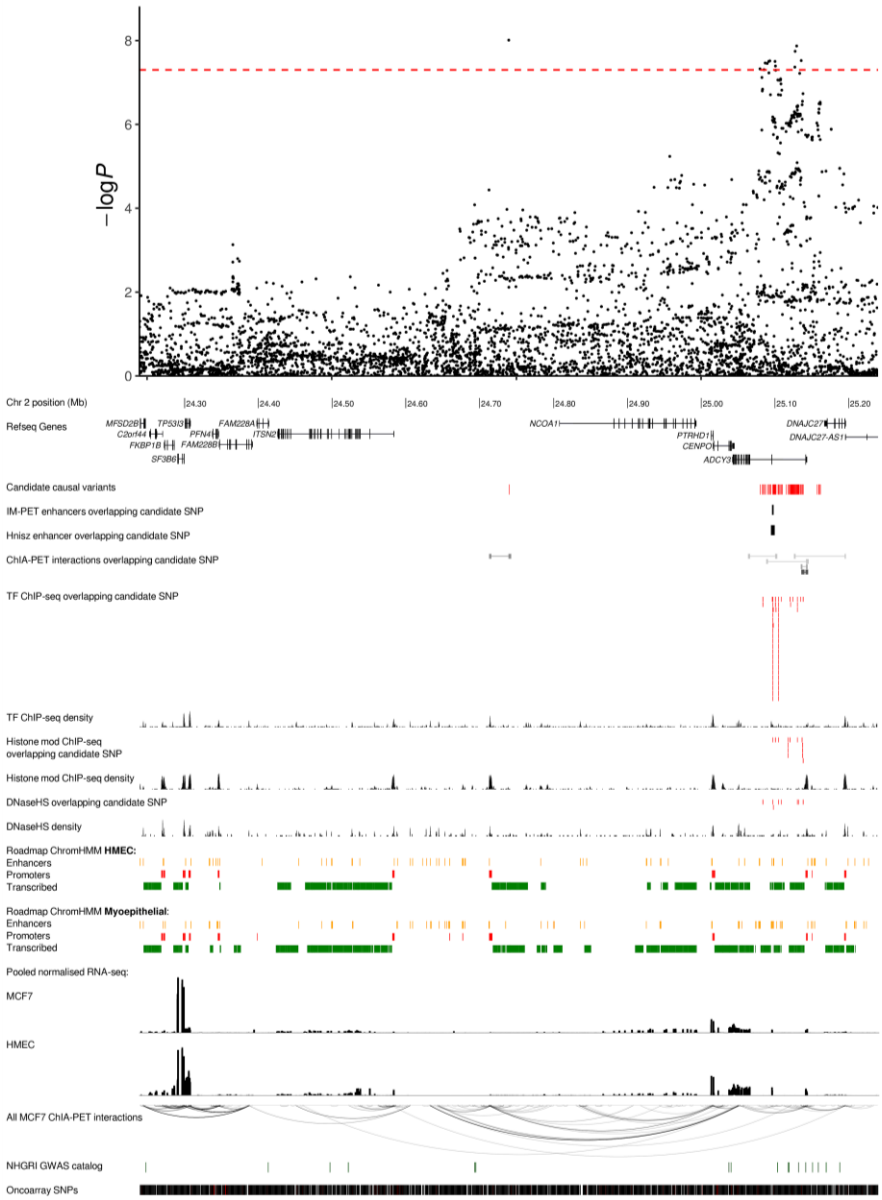
Supplementary Figure 2

Manhattan plot of associations with breast cancer risk for all imputed and genotyped SNPs using combined data from ER-negative cases and controls and *BRCA1* mutation carriers, after excluding known breast cancer susceptibility loci.



Supplementary Figure 3

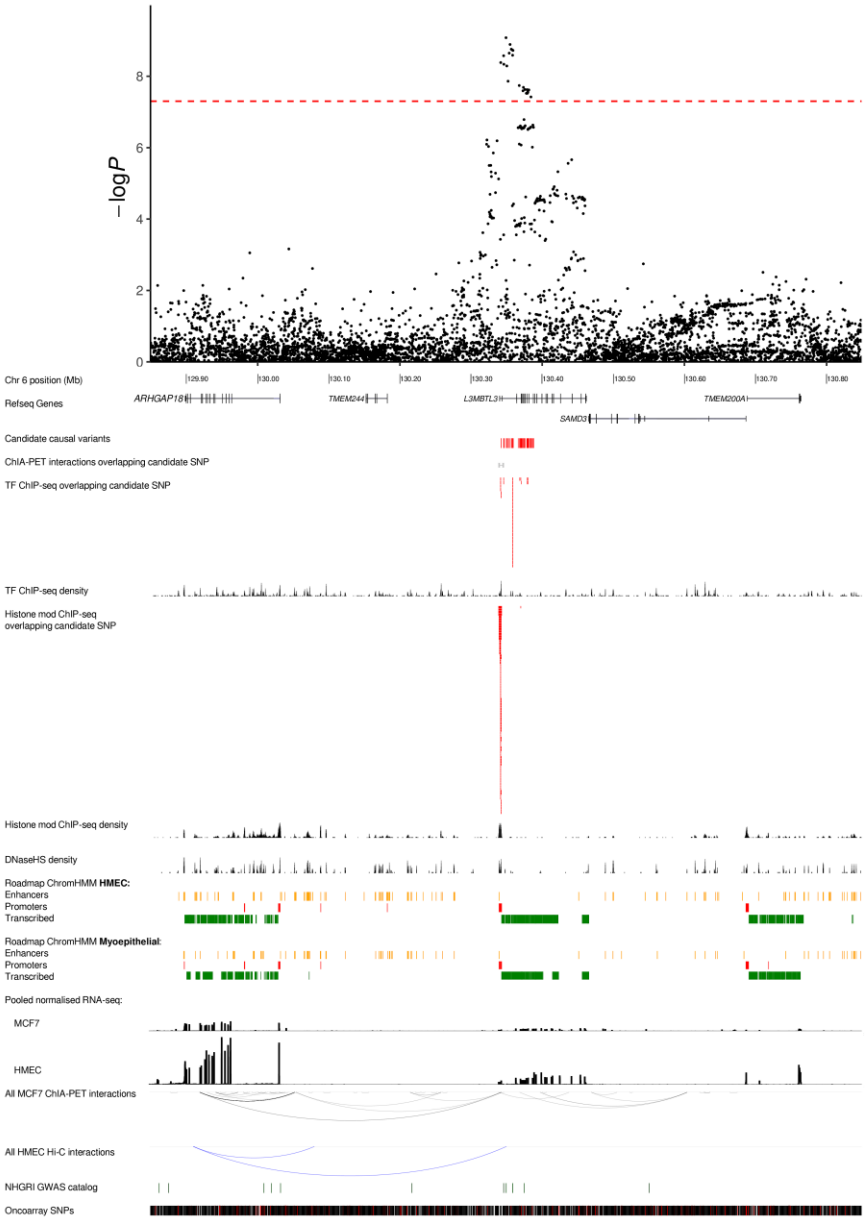
Quantile–quantile plot of associations with breast cancer risk for all imputed and genotyped SNPs using combined data from ER-negative cases and controls and *BRCA1* mutation carriers.



Supplementary Figure 4

Genomic region around the ER-negative risk-associated variant 2_24739694_CT_T (rs200648189).

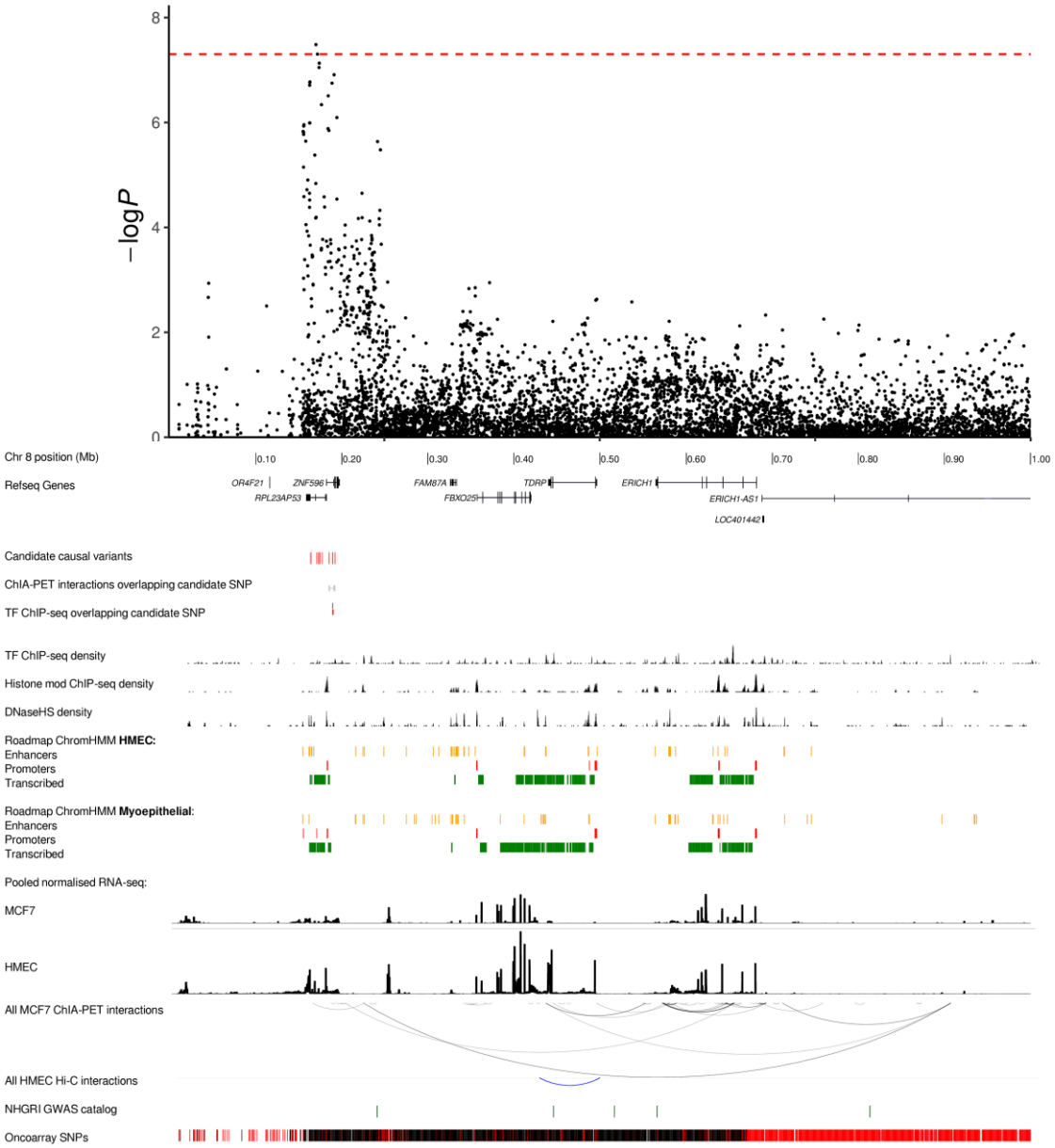
One-megabase region showing the statistical significance of all genotyped and imputed SNPs (regional Manhattan plot) and the positions of candidate causal variants in relation to RefSeq annotated genes. Breast cell enhancers overlapping candidate SNPs predicted to target nearby genes by methods including IM-PET and Hnisz are depicted as black bars. Chromatin interactions from ENCODE ChIA-PET experiments in MCF7 cells overlapping candidate variants are depicted as boxes connected by thin lines and are shaded to reflect the confidence score of the interaction. Epigenomic features (derived from publicly available transcription factor ChIP–seq, histone modification ChIP–seq and DNase–seq) that overlap candidate variants are shown as red segments. Density tracks show the summed occurrence of transcription factor ChIP–seq, histone modification ChIP–seq, and DHS peaks at each genomic position. Roadmap Epigenomics Project chromatin state models for HMECs and myoepithelial cells grouped into enhancer, promoter or transcribed annotations are shown as yellow, red and green segments, respectively. Transcript levels in MCF7 cells and HMECs are represented by histograms depicting the mean of combined and normalized RNA–seq expression levels at each genomic position. All MCF7 ChIA-PET (ENCODE) chromatin interactions are represented by black arcs. Published GWAS signals from the NHGRI catalog are shown as green ticks. The last track shows tested OncoArray SNPs (genotyped or imputed) as black ticks and uninterrogated, common (dbSNP138 EUR MAF > 1%) SNPs as red ticks.



Supplementary Figure 5

Genomic region around the ER-negative risk-associated variant 6_130349119_T_C (rs6569648).

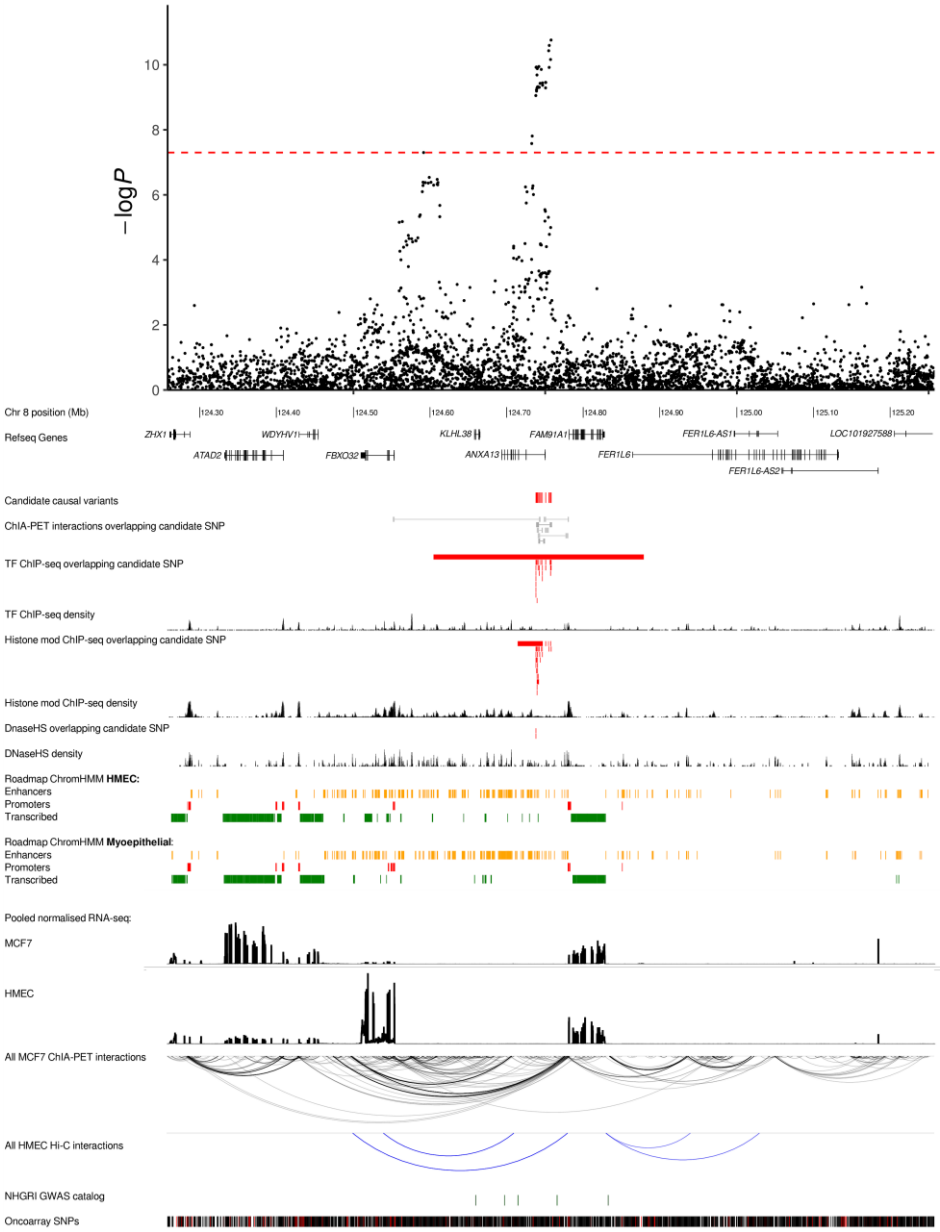
One-megabase region showing the statistical significance of all genotyped and imputed SNPs (regional Manhattan plot) and the positions of candidate causal variants in relation to RefSeq annotated genes. Chromatin interactions from ENCODE ChIA-PET experiments in MCF7 cells overlapping candidate variants are depicted as boxes connected by thin lines and are shaded to reflect the confidence score of the interaction. Epigenomic features (derived from publicly available transcription factor ChIP–seq and histone modification ChIP–seq) that overlap candidate variants are shown as red segments. Density tracks show the summed occurrence of transcription factor ChIP–seq, histone modification ChIP–seq and DHS peaks at each genomic position. Roadmap Epigenomics Project chromatin state models for HMECs and myoepithelial cells grouped into enhancer, promoter or transcribed annotations are shown as yellow, red and green segments, respectively. Transcript levels in MCF7 cells and HMECs are represented by histograms depicting the mean of combined and normalized RNA–seq expression levels at each genomic position. All MCF7 ChIA-PET (ENCODE) and HMEC Hi-C[52] chromatin interactions are represented by black and blue arcs, respectively. Published GWAS signals from the NHGRI GWAS catalog are shown as green ticks. The last track shows tested OncoArray SNPs (genotyped or imputed) as black ticks and uninterrogated, common (dbSNP138 EUR MAF > 1%) SNPs as red ticks.



Supplementary Figure 6

Genomic region around the ER-negative risk-associated variant 8_170692_T_C (rs66823261).

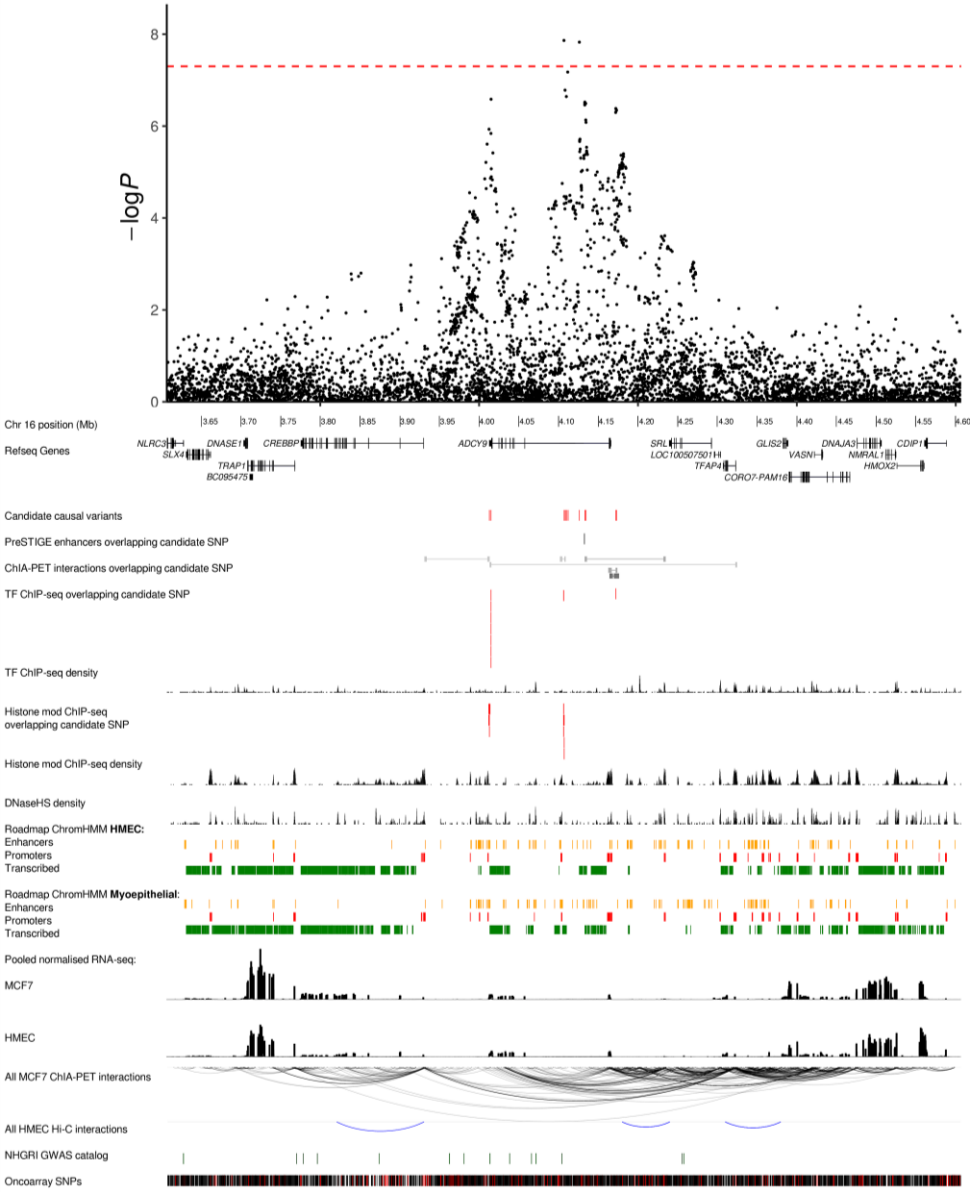
One-megabase region showing the statistical significance of all genotyped and imputed SNPs (regional Manhattan plot) and the positions of candidate causal variants in relation to RefSeq annotated genes. Chromatin interactions from ENCODE ChIA-PET experiments in MCF7 cells overlapping candidate variants are depicted as boxes connected by thin lines and are shaded to reflect the confidence score of the interaction. Epigenomic features derived from publicly available transcription factor ChIP–seq that overlap candidate variants are shown as red segments. Density tracks show the summed occurrence of transcription factor ChIP–seq, histone modification ChIP–seq and DHS peaks at each genomic position. Roadmap Epigenomics Project chromatin state models for HMECs and myoepithelial cells grouped into enhancer, promoter or transcribed annotations are shown as yellow, red and green segments, respectively. Transcript levels in MCF7 cells and HMECs are represented by histograms depicting the mean of combined and normalized RNA–seq expression levels at each genomic position. All MCF7 ChIA-PET (ENCODE) and HMEC Hi-C[52] chromatin interactions are represented by black and blue arcs, respectively. Published GWAS signals from the NHGRI GWAS catalog are shown as green ticks. The last track shows tested OncoArray SNPs (genotyped or imputed) as black ticks and uninterrogated, common (dbSNP138 EUR MAF > 1%) SNPs as red ticks.



Supplementary Figure 7

Genomic region around the ER-negative risk-associated variant 8_124757661_C_T (rs17350191).

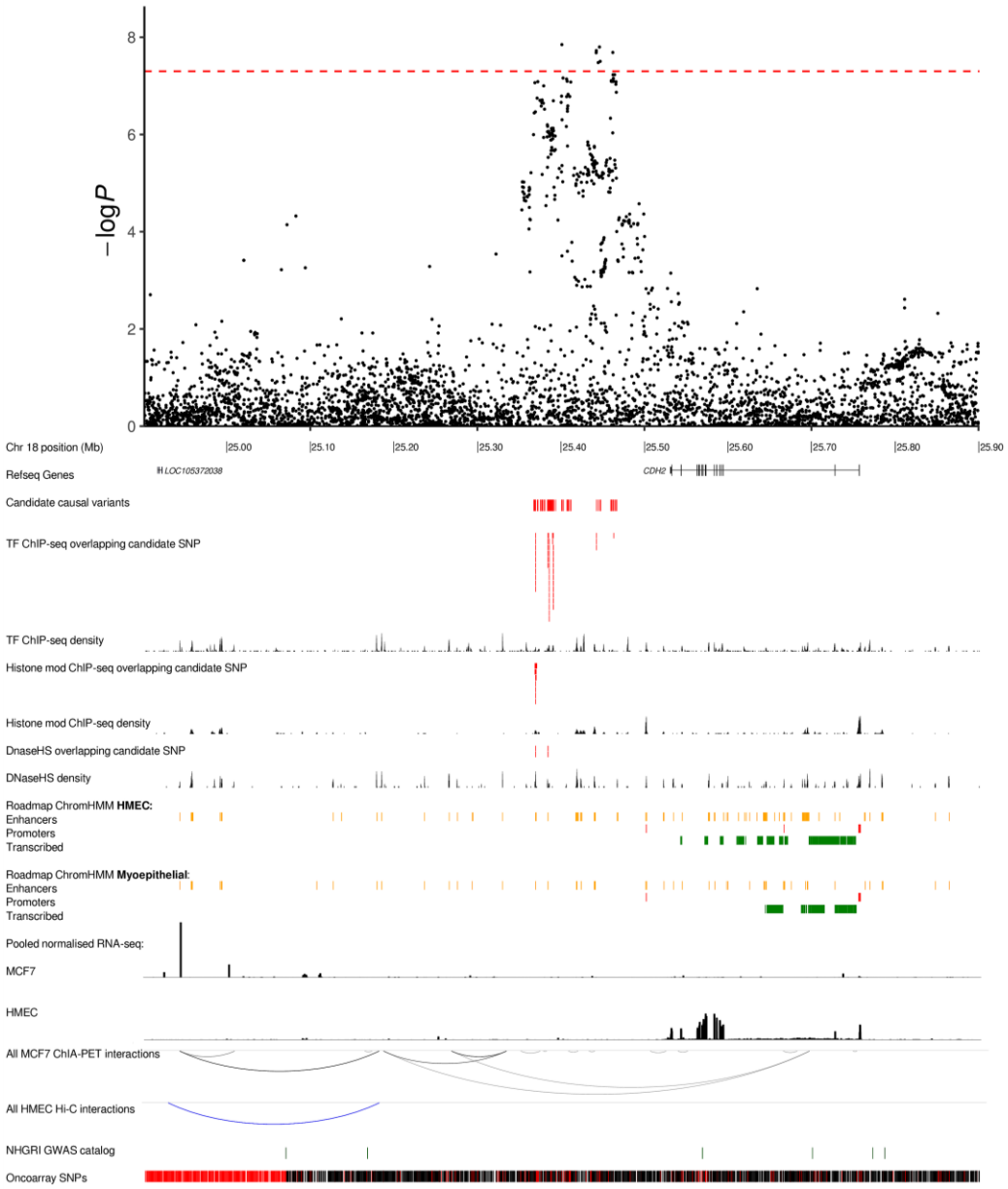
One-megabase region showing the statistical significance of all genotyped and imputed SNPs (regional Manhattan plot) and the positions of candidate causal variants in relation to RefSeq annotated genes. Chromatin interactions from ENCODE ChIA-PET experiments in MCF7 cells overlapping candidate variants are depicted as boxes connected by thin lines and are shaded to reflect the confidence score of the interaction. Epigenomic features (derived from publicly available transcription factor ChIP–seq, histone modification ChIP–seq and DNase–seq) that overlap candidate variants are shown as red segments. Density tracks show the summed occurrence of transcription factor ChIP–seq, histone modification ChIP–seq and DHS peaks at each genomic position. Roadmap Epigenomics Project chromatin state models for HMECs and myoepithelial cells grouped into enhancer, promoter or transcribed annotations are shown as yellow, red and green segments, respectively. Transcript levels in MCF7 cells and HMECs are represented by histograms depicting the mean of combined and normalized RNA–seq expression levels at each genomic position. All MCF7 ChIA-PET (ENCODE) and HMEC Hi-C[52] chromatin interactions are represented by black and blue arcs, respectively. Published GWAS signals from the NHGRI GWAS catalog are shown as green ticks. The last track shows tested OncoArray SNPs (genotyped or imputed) as black ticks and uninterrogated, common (dbSNP138 EUR MAF > 1%) SNPs as red ticks.



Supplementary Figure 8

Genomic region around the ER-negative risk-associated variant 16_4106788_C_A (rs11076805).

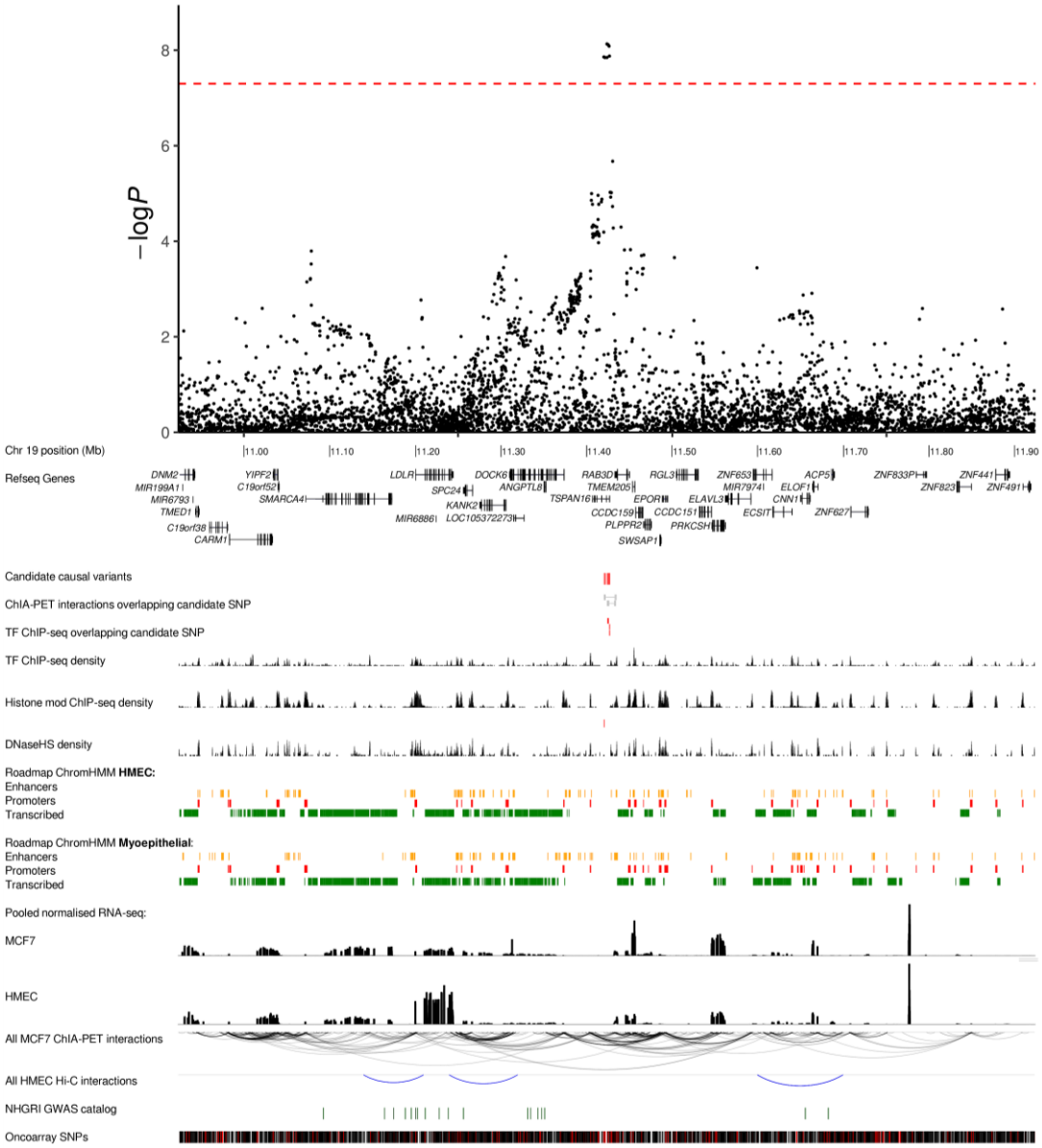
One-megabase region showing the statistical significance of all genotyped and imputed SNPs (regional Manhattan plot) and the positions of candidate causal variants in relation to RefSeq annotated genes. Breast cell enhancers overlapping candidate SNPs predicted to target nearby genes by PreSTIGE[53] are depicted as black bars. Chromatin interactions from ENCODE ChIA-PET experiments in MCF7 cells overlapping candidate variants are depicted as boxes connected by thin lines and are shaded to reflect the confidence score of the interaction. Epigenomic features (derived from publicly available transcription factor ChIP–seq and histone modification ChIP–seq) that overlap candidate variants are shown as red segments. Density tracks show the summed occurrence of transcription factor ChIP–seq, histone modification ChIP–seq and DHS peaks at each genomic position. Roadmap Epigenomics Project chromatin state models for HMECs and myoepithelial cells grouped into enhancer, promoter or transcribed annotations are shown as yellow, red and green segments, respectively. Transcript levels in MCF7 cells and HMECs are represented by histograms depicting the mean of combined and normalized RNA–seq expression levels at each genomic position. All MCF7 ChIA-PET (ENCODE) and HMEC Hi-C[52] chromatin interactions are represented by black and blue arcs, respectively. Published GWAS signals from the NHGRI GWAS catalog are shown as green ticks. The last track shows tested OncoArray SNPs (genotyped or imputed) as black ticks and uninterrogated, common (dbSNP138 EUR MAF > 1%) SNPs as red ticks.



Supplementary Figure 9

Genomic region around the ER-negative risk-associated variant 18_25401204_A_AT (rs36194942).

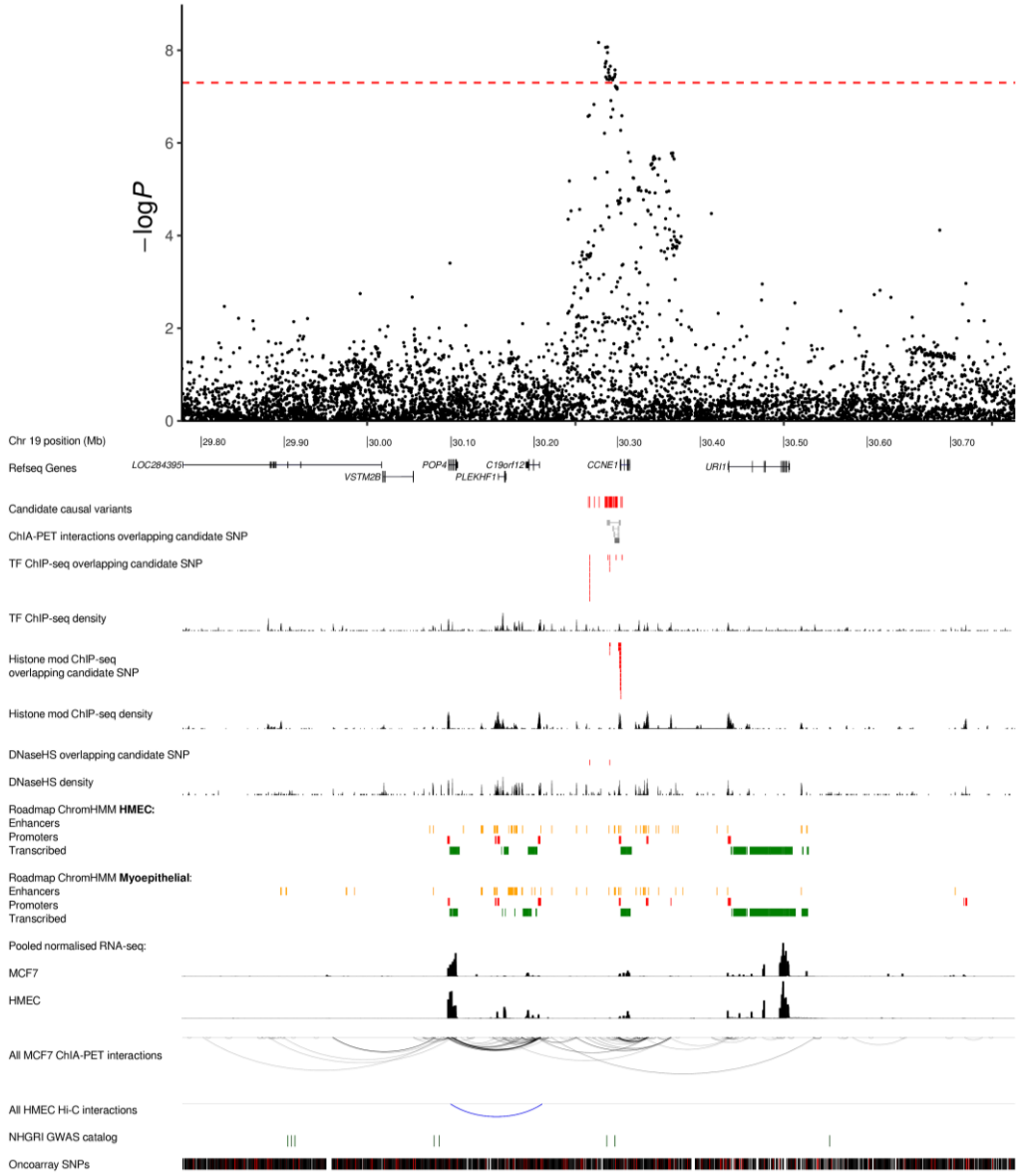
One-megabase region showing the statistical significance of all genotyped and imputed SNPs (regional Manhattan plot) and the positions of candidate causal variants in relation to RefSeq annotated genes. Epigenomic features (derived from publicly available transcription factor ChIP–seq, histone modification ChIP–seq and DNase–seq) that overlap candidate variants are shown as red segments. Density tracks show the summed occurrence of transcription factor ChIP–seq, histone modification ChIP–seq and DHS peaks at each genomic position. Roadmap Epigenomics Project chromatin state models for HMECs and myoepithelial cells grouped into enhancer, promoter or transcribed annotations are shown as yellow, red and green segments, respectively. Transcript levels in MCF7 cells and HMECs are represented by histograms depicting the mean of combined and normalized RNA–seq expression levels at each genomic position. All MCF7 ChIA-PET (ENCODE) and HMEC Hi-C[52] chromatin interactions are represented by black and blue arcs, respectively. Published GWAS signals from the NHGRI GWAS catalog are shown as green ticks. The last track shows tested OncoArray SNPs (genotyped or imputed) as black ticks and uninterrogated, common (dbSNP138 EUR MAF > 1%) SNPs as red ticks.



Supplementary Figure 10

Genomic region around the ER-negative risk-associated variant 19_11423703_C_G (rs322144).

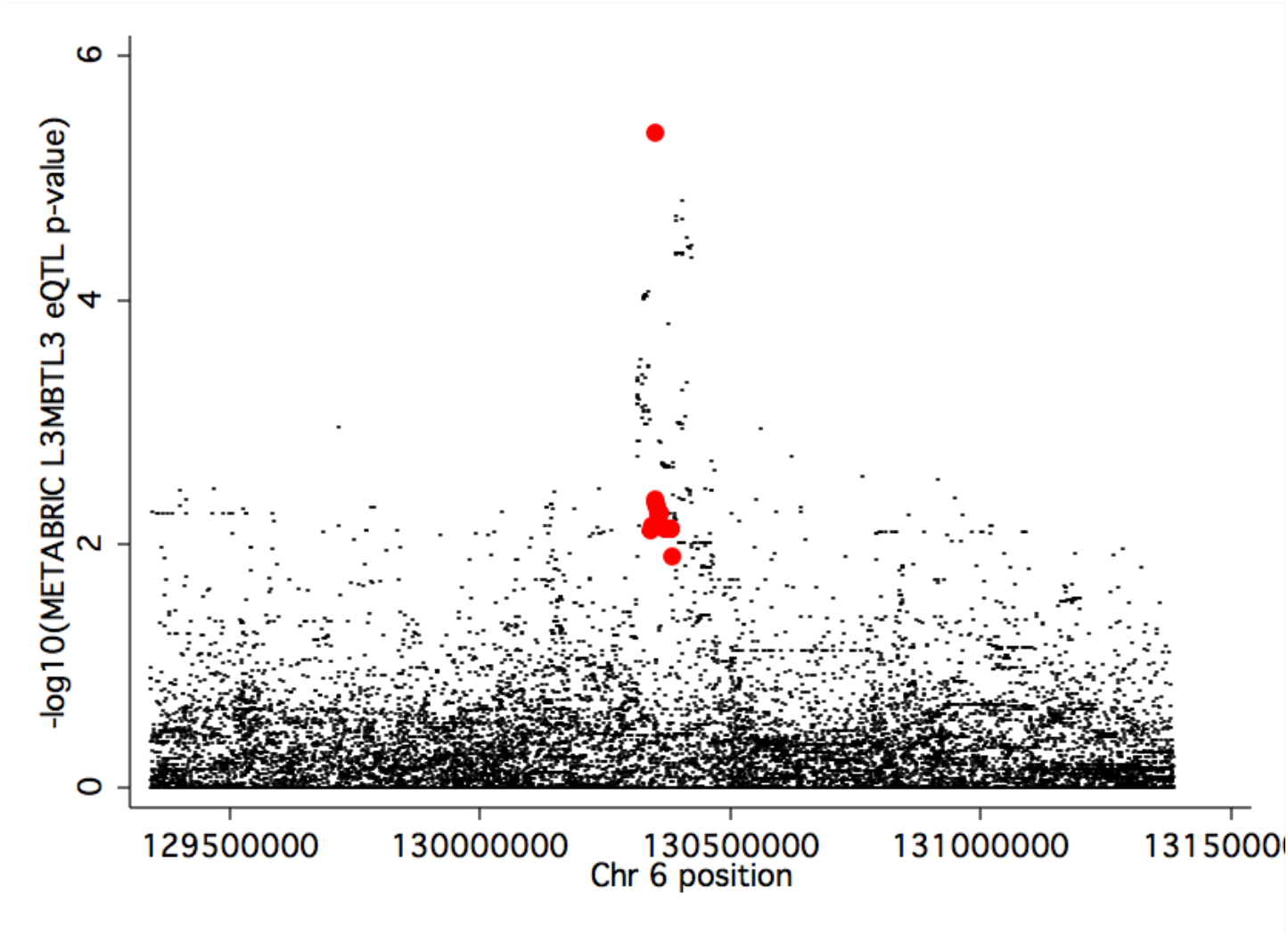
One-megabase region showing the statistical significance of all genotyped and imputed SNPs (regional Manhattan plot) and the positions of candidate causal variants in relation to RefSeq annotated genes. Chromatin interactions from ENCODE ChIA-PET experiments in MCF7 cells overlapping candidate variants are depicted as boxes connected by thin lines and are shaded to reflect the confidence score of the interaction. Epigenomic features (derived from publicly available transcription factor ChIP–seq and DNase–seq) that overlap candidate variants are shown as red segments. Density tracks show the summed occurrence of transcription factor ChIP–seq, histone modification ChIP–seq and DHS peaks at each genomic position. Roadmap Epigenomics Project chromatin state models for HMECs and myoepithelial cells grouped into enhancer, promoter or transcribed annotations are shown as yellow, red and green segments, respectively. Transcript levels in MCF7 cells and HMECs are represented by histograms depicting the mean of combined and normalized RNA–seq expression levels at each genomic position. All MCF7 ChIA-PET (ENCODE) and HMEC Hi-C[52] chromatin interactions are represented by black and blue arcs, respectively. Published GWAS signals from the NHGRI GWAS catalog are shown as green ticks. The last track shows tested OncoArray SNPs (genotyped or imputed) as black ticks and uninterrogated, common (dbSNP138 EUR MAF > 1%) SNPs as red ticks.



Supplementary Figure 11

Genomic region around the ER-negative risk-associated variant 19_30277729_C_T (rs113701136).

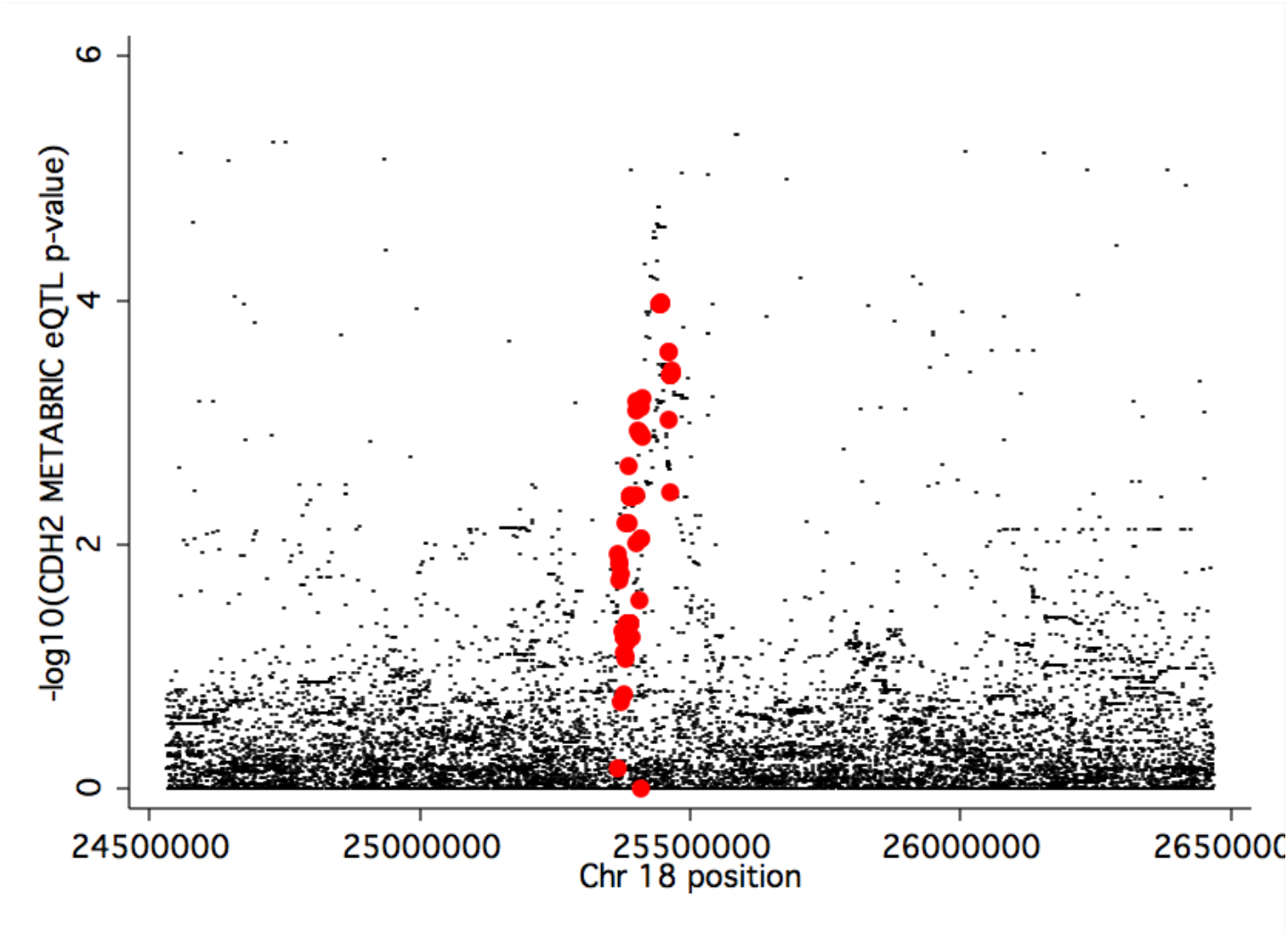
One-megabase region showing the statistical significance of all genotyped and imputed SNPs (regional Manhattan plot) and the positions of candidate causal variants in relation to RefSeq annotated genes. Chromatin interactions from ENCODE ChIA-PET experiments in MCF7 cells overlapping candidate variants are depicted as boxes connected by thin lines and are shaded to reflect the confidence score of the interaction. Epigenomic features (derived from publicly available transcription factor ChIP–seq, histone modification ChIP–seq and DNase–seq) that overlap candidate variants are shown as red segments. Density tracks show the summed occurrence of transcription factor ChIP–seq, histone modification ChIP–seq and DHS peaks at each genomic position. Roadmap Epigenomics Project chromatin state models for HMECs and myoepithelial cells grouped into enhancer, promoter or transcribed annotations are shown as yellow, red and green segments, respectively. Transcript levels in MCF7 cells and HMECs are represented by histograms depicting the mean of combined and normalized RNA–seq expression levels at each genomic position. All MCF7 ChIA-PET (ENCODE) and HMEC Hi-C[52] chromatin interactions are represented by black and blue arcs, respectively. Published GWAS signals from the NHGRI GWAS catalog are shown as green ticks. The last track shows tested OncoArray SNPs (genotyped or imputed) as black ticks and uninterrogated, common (dbSNP138 EUR MAF > 1%) SNPs as red ticks.



Supplementary Figure 12

Regional eQTL association plot for all variants within 1 Mb of *L3MTBL3* and expression of *L3MTBL3*.

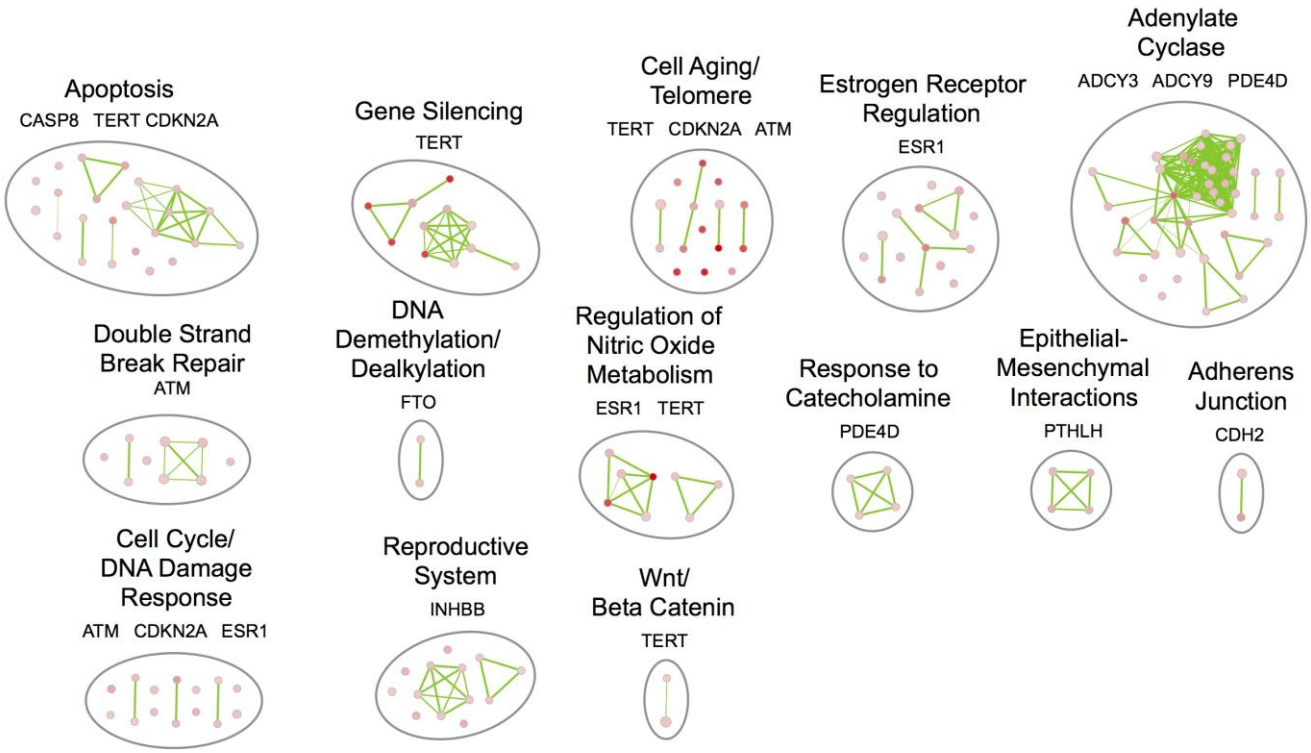
Red dots indicate candidate causal risk variants from the meta-analysis of BCAC ER-negative case–control and CIMBA *BRCA1* mutation carrier data.



Supplementary Figure 13

Regional eQTL association plot for all variants within 1 Mb of *CDH2* and expression of *CDH2*.

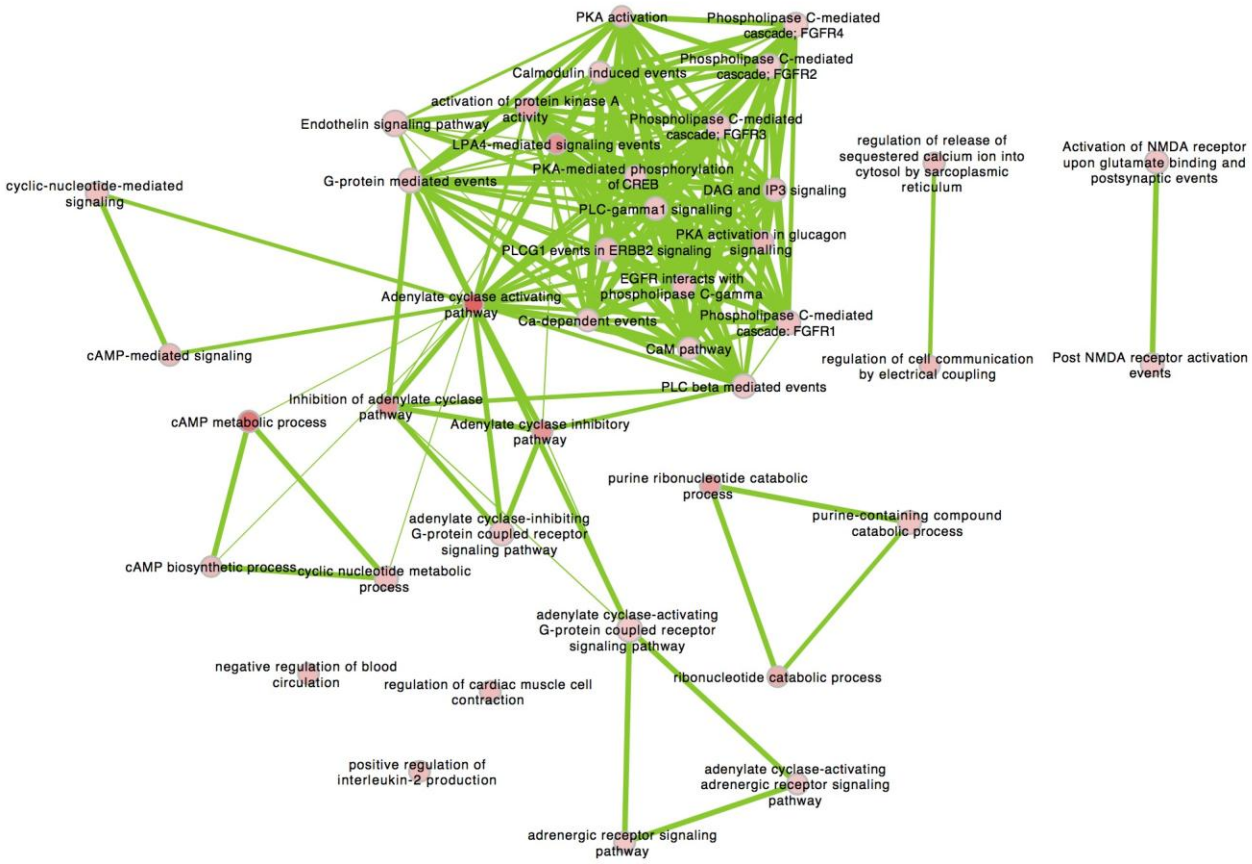
Red dots indicate candidate causal risk variants from the meta-analysis of BCAC ER-negative case–control and CIMBA *BRCA1* mutation carrier data.



Supplementary Figure 14

Enrichment map for pathways enriched in susceptibility to ER-negative breast cancer.

Enriched pathways (enrichment score (ES) ≥ 0.41) are grouped into themes and annotated with genes that appeared to drive the enrichment signal (Online Methods). Shaded circles represent pathways (darker red indicates higher ES, and larger size denotes a greater number of genes in the pathway), and green lines connect those that are most similar in terms of gene set overlap (>70%), with thicker lines denoting greater similarity.



Supplementary Figure 15

Enrichment map of the adenylate cyclase theme, enriched in susceptibility to ER-negative breast cancer.

Shaded circles represent pathways (darker red indicates higher ES, and larger size denotes a greater number of genes in the pathway), and green lines connect those that are most similar in terms of gene set overlap (>70%), with thicker lines denoting greater similarity.

Members of consortia listed as authors

ABCTB Investigators

Christine Clarke (Westmead Institute for Medical Research, University of Sydney, NSW, Australia); Rosemary Balleine (Pathology West ICPMR, Westmead, NSW, Australia); Robert Baxter (Kolling Institute of Medical Research, University of Sydney, Royal North Shore Hospital, NSW, Australia); Stephen Braye (Pathology North, John Hunter Hospital, Newcastle, NSW, 2305, Australia); Jane Carpenter (Westmead Institute for Medical Research, University of Sydney); Jane Dahlstrom (Department of Anatomical Pathology, ACT Pathology, Canberra Hospital, ACT, Australia; ANU Medical School, Australian National University, ACT, Australia); John Forbes (Department of Surgical Oncology, Calvary Mater Newcastle Hospital, Australian New Zealand Breast Cancer Trials Group, and School of Medicine and Public Health, University of Newcastle, NSW, Australia); C Soon Lee (School of Science and Health, The University of Western Sydney, Sydney, Australia); Deborah Marsh (Hormones and Cancer Group, Kolling Institute of Medical Research, Royal North Shore Hospital, University of Sydney, NSW, Australia); Adrienne Morey (SydPath St Vincent's Hospital, Sydney, NSW, Australia); Nirmala Pathmanathan (Department of Tissue Pathology and Diagnostic Oncology, Pathology West; Westmead Breast Cancer Institute, Westmead Hospital, NSW, Australia); Rodney Scott (Centre for Information Based Medicine, Hunter Medical Research Institute, NSW, 2305, Australia; Priority Research Centre for Cancer, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, NSW, Australia); Peter Simpson (The University of Queensland: UQ Centre for Clinical Research and School of Medicine, QLD, Australia); Allan Spigelman (Hereditary Cancer Clinic, St Vincent's Hospital, The Kinghorn Cancer Centre, Sydney, New South Wales, 2010, Australia); Nicholas Wilcken (Crown Princess Mary Cancer Centre, Westmead Hospital, Westmead, Australia; Sydney Medical School - Westmead, University of Sydney, NSW, Australia); Desmond Yip (Department of Medical Oncology, The Canberra Hospital, ACT, Australia; ANU Medical School, Australian National University, ACT, Australia); Nikolajs Zeps (St John of God Perth Northern Hospitals, Perth, WA, Australia)

EMBRACE

Helen Gregory (North of Scotland Regional Genetics Service, NHS Grampian & University of Aberdeen, Foresterhill, Aberdeen, UK); Zosia Miedzybrodzka (North of Scotland Regional Genetics Service, NHS Grampian & University of Aberdeen, Foresterhill, Aberdeen, UK); Patrick J. Morrison (Northern Ireland Regional Genetics Centre, Belfast Health and Social Care Trust, and Department of Medical Genetics, Queens University Belfast, Belfast, UK); Kai-ren Ong (West Midlands Regional Genetics Service, Birmingham Women’s Hospital Healthcare NHS Trust, Edgbaston, Birmingham, UK); Alan Donaldson (Clinical Genetics Department, St Michael’s Hospital, Bristol, UK); Marc Tischkowitz (Department of Medical Genetics, University of Cambridge, UK); Mark T. Rogers (All Wales Medical Genetics Services, University Hospital of Wales, Cardiff, UK); M. John Kennedy (Academic Unit of Clinical and Molecular Oncology, Trinity College Dublin and St James's Hospital, Dublin, Eire); Mary E. Porteous (South East of Scotland Regional Genetics Service, Western General Hospital, Edinburgh, UK); Carole Brewer (Department of Clinical Genetics, Royal Devon & Exeter Hospital, Exeter, UK); Rosemarie Davidson (Clinical Genetics, Southern General Hospital, Glasgow, UK); Louise Izatt (Clinical Genetics, Guy’s and St. Thomas’ NHS Foundation Trust, London, UK); Angela Brady (North West Thames Regional

Genetics Service, Kennedy-Galton Centre, Harrow, UK); Julian Barwell (Leicestershire Clinical Genetics Service, University Hospitals of Leicester NHS Trust, UK); Julian Adlard (Yorkshire Regional Genetics Service, Leeds, UK); Claire Foo (Department of Clinical Genetics, Alder Hey Hospital, Eaton Road, Liverpool, UK); D. Gareth Evans (Genetic Medicine, Manchester Academic Health Sciences Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK); Fiona Lalloo (Genetic Medicine, Manchester Academic Health Sciences Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK); Lucy E. Side (North East Thames Regional Genetics Service, Great Ormond Street Hospital for Children NHS Trust, London, UK); Jacqueline Eason (Nottingham Clinical Genetics Service, Nottingham University Hospitals NHS Trust, UK); Alex Henderson (Institute of Genetic Medicine, Centre for Life, Newcastle Upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, UK); Lisa Walker (Oxford Regional Genetics Service, Churchill Hospital, Oxford, UK); Rosalind A. Eeles (Oncogenetics Team, The Institute of Cancer Research and Royal Marsden NHS Foundation Trust, UK); Jackie Cook (Sheffield Clinical Genetics Service, Sheffield Children's Hospital, Sheffield, UK); Katie Snape (South West Thames Regional Genetics Service, St.Georges Hospital, Cranmer Terrace, Tooting, London, UK); Diana Eccles (University of Southampton Faculty of Medicine, Southampton University Hospitals NHS Trust, Southampton, UK); Alex Murray (All Wales Medical Genetics Services, Singleton Hospital, Swansea, UK); Emma McCann (All Wales Medical Genetics Service, Glan Clwyd Hospital, Rhyl, UK).

GEMO Study Collaborators

Dominique Stoppa-Lyonnet, Muriel Belotti, Anne-Marie Birot, Bruno Buecher, Emmanuelle Fourme, Marion Gauthier-Villars, Lisa Golmard, Claude Houdayer, Virginie Moncoutier, Antoine de Pauw, Claire Saule (Service de Génétique, Institut Curie, Paris, France); Fabienne Lesueur, Noura Mebirouk (Inserm U900, Institut Curie, Paris, France); Olga Sinilnikova†, Sylvie Mazoyer, Francesca Damiola, Laure Barjhoux, Carole Verny-Pierre, Mélanie Léone, Nadia Boutry-Kryza, Alain Calender, Sophie Giraud (Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon - Centre Léon Bérard, Lyon, France); Olivier Caron, Marine Guillaud-Bataille (Institut Gustave Roussy, Villejuif, France: Brigitte Bressac-de-Paillerets); Yves-Jean Bignon, Nancy Uhrhammer (Centre Jean Perrin, Clermont–Ferrand, France); Christine Lasset, Valérie Bonadona (Centre Léon Bérard, Lyon, France); Pascaline Berthet, Dominique Vaur, Laurent Castera (Centre François Baclesse, Caen, France); Hagay Sobol, Violaine Bourdon, Tetsuro Noguchi, Audrey Remenieras, François Eisinger, Catherine Noguès (Institut Paoli Calmettes, Marseille, France); Isabelle Coupier, Pascal Pujol (CHU Arnaud-de-Villeneuve, Montpellier, France); Jean-Philippe Peyrat, Joëlle Fournier, Françoise Révillion, Claude Adenis (Centre Oscar Lambret, Lille, France); Danièle Muller, Jean-Pierre Fricker (Centre Paul Strauss, Strasbourg, France); Emmanuelle Barouk-Simonet, Françoise Bonnet, Virginie Bubien, Nicolas Sevenet, Michel Longy (Institut Bergonié, Bordeaux, France); Christine Toulas, Rosine Guimbaud, Laurence Gladieff, Viviane Feillel (Institut Claudius Regaud, Toulouse, France); Dominique Leroux, Hélène Dreyfus, Christine Rebischung, Magalie Peysselon (CHU Grenoble, France); Fanny Coron, Laurence Faivre, Amandine Baurand, Caroline Jacquot, Geoffrey, Bertolone, Sarab Lizard (CHU Dijon, France); Fabienne Prieur, Marine Lebrun, Caroline Kientz (CHU St-Etienne, France); Sandra Fert Ferrer (Hôtel Dieu Centre Hospitalier, Chambéry, France); Véronique Mari (Centre Antoine Lacassagne, Nice, France); Laurence Vénat-Bouvet (CHU Limoges, France); Capucine Delnatte, Stéphane Bézieau (CHU Nantes, France); Isabelle Mortemousque (CHU

Bretonneau, Tours and Centre Hospitalier de Bourges France); Florence Coulet, Chrystelle Colas, Florent Soubrier, Mathilde Warcoin (Groupe Hospitalier Pitié-Salpétrière, Paris, France); Johanna Sokolowska, Myriam Bronner (CHU Vandoeuvre-les-Nancy, France); Marie-Agnès Collonge-Rame, Alexandre Damette (CHU Besançon, France); Paul Gesta (CHU Poitiers, Centre Hospitalier d'Angoulême and Centre Hospitalier de Niort, France); Hakima Lallaoui (Centre Hospitalier de La Rochelle); Jean Chiesa (CHU Nîmes Carémeau, France); Denise Molina-Gomes (CHI Poissy, France); Olivier Ingster (CHU Angers, France).

HEBON

Annemieke H. van der Hout (Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands); Ans M.W. van den Ouweland (Department of Clinical Genetics, Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, The Netherlands); Arjen R. Mensenkamp (Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands); Carolien H.M. van Deurzen (Department of Pathology, Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, The Netherlands); Carolien M. Kets (Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands); Caroline Seynaeve (Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, The Netherlands); Christi J. van Asperen (Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands); Cora M. Aalfs (Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands); Encarna B. Gómez Garcia (Department of Clinical Genetics , MUMC, Maastricht, The Netherlands); Flora E. van Leeuwen (Department of Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands); Frans B.L. Hogervorst (Family Cancer Clinic, Netherlands Cancer Institute, Amsterdam, The Netherlands); G.H. de Bock (Department of Oncological Epidemiology, University Medical Center, Groningen University, Groningen, The Netherlands); Hanne E.J. Meijers-Heijboer (Department of Clinical Genetics, VU University Medical Centre, Amsterdam, The Netherlands); Inge M. Obdeijn (Department of Radiology, Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, The Netherlands); J. Margriet Collée (Department of Clinical Genetics, Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, The Netherlands); J.J.P. Gille (Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands); Jan C. Oosterwijk (Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands); Juul T. Wijnen (Department of Human Genetics & Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands); Lizet E. van der Kolk (Family Cancer Clinic, Netherlands Cancer Institute, Amsterdam, The Netherlands); Maartje J. Hooning (Department of Medical Oncology, Family Cancer Clinic, Erasmus MC Cancer Institute, Rotterdam, The Netherlands); Margreet G.E.M. Ausems (Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands); Marian J.E. Mourits (Department of Gynaecological Oncology, University Medical Center, Groningen University, Groningen, The Netherlands); Marinus J. Blok (Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands); Marjanka K. Schmidt (Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands); Matti A. Rookus (Department of Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands); Muriel A. Adank (Department of Clinical Genetics, VU University Medical Centre, Amsterdam, The Netherlands); Peter Devilee (Department of Human Genetics & Department of Pathology, Leiden University Medical Center, Leiden, The

Netherlands); Rob B. van der Luijt (Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands); T.C.T.E.F. van Cronenburg (Department of Clinical Genetics Leiden University Medical Center Leiden, The Netherlands); Carmen C. van der Pol (Department of Oncological and Endocrine Surgery, University Medical Center Utrecht, Utrecht, The Netherlands):Nicola S. Russell (Department of Radiotherapy, Netherlands Cancer Institute, Amsterdam, The Netherlands); Sabine Siesling (The Netherlands Comprehensive Cancer Organization (IKNL), Utrecht, The Netherlands); Lucy Overbeek (The nationwide network and registry of histo- and cytopathology in The Netherlands [PALGA]); R. Wijnands (Department of Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands); Judith L. de Lange (Department of Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands).

kConFab/AOCS Investigators

Stephen Fox, Ian Campbell (Peter MacCallum Cancer Centre, Melbourne, Australia); Georgia Chenevix-Trench, Amanda Spurdle, Penny Webb (QIMR Berghofer Medical Research Institute, Brisbane, Australia); Anna de Fazio (Westmead Millenium Institute, Sydney, Australia); Margaret Tassell (BCNA delegate, Community Representative); Judy Kirk (Westmead Hospital, Sydney, Australia); Geoff Lindeman (Walter and Eliza Hall Institute, Melbourne, Australia); Melanie Price (University of Sydney, Sydney, Australia); Melissa Southey (University of Melbourne, Melbourne, Australia); Roger Milne (Cancer Council Victoria, Melbourne, Australia); Sid Deb (Melbourne Health, Melbourne, Australia); David Bowtell (Garvan Institute of Medical Research, Sydney, Australia).

NBSC Collaborators

Kristine K. Sahlberg (Department of Research, Vestre Viken Hospital, Drammen, Norway; Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway); Lars Ottestad (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway); Rolf Kåresen (Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; Department of Breast- and Endocrine Surgery, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital Ullevål, Oslo, Norway); Anita Langerød (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway); Ellen Schlichting (Department of Breast- and Endocrine Surgery, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital Ullevål, Oslo, Norway); Marit Muri Holmen (Department of Radiology and Nuclear Medicine, Oslo University Hospital, Oslo, Norway); Toril Sauer (Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; Department of Pathology at Akershus University hospital, Lørenskog, Norway); Vilde Haakensen (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway); Olav Engebråten (Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway; Department of Oncology, Division of Surgery and Cancer and Transplantation Medicine, Oslo University Hospital Radiumhospitalet, Oslo, Norway); Bjørn Naume (Department of Oncology, Division of Surgery and Cancer and Transplantation Medicine, Oslo University Hospital Radiumhospitalet, Oslo, Norway); Cecile E. Kiserud (National Advisory Unit on Late Effects after

Cancer Treatment, Department of Oncology, Oslo University Hospital, Oslo, Norway; Department of Oncology, Oslo University Hospital Ullevål, Oslo, Norway); Kristin V. Reinertsen (National Advisory Unit on Late Effects after Cancer Treatment, Department of Oncology, Oslo University Hospital, Oslo, Norway; Department of Oncology, Oslo University Hospital Ullevål, Oslo, Norway); Åslaug Helland (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway; Department of Oncology, Division of Surgery and Cancer and Transplantation Medicine, Oslo University Hospital Radiumhospitalet, Oslo, Norway); Margit Riis (Department of Breast- and Endocrine Surgery, Division of Surgery, Cancer and Transplantation Medicine, Oslo University Hospital Ullevål, Oslo, Norway); Jürgen Geisler (Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; Department of Breast-Endocrine Surgery, Akershus University Hospital, Oslo, Norway); Per Eystein Lønning (Section of Oncology, Institute of Medicine, University of Bergen and Department of Oncology, Haukeland University Hospital, Bergen, Norway); Solveig Hofvind (Cancer Registry of Norway, Oslo, Norway, Oslo; Akershus University College of Applied Sciences, Faculty of Health Science, Oslo, Norway); Tone F. Bathen (Department of Circulation and Medical Imaging, Norwegian University of Science and Technology (NTNU), Trondheim, Norway); Elin Borgen (Department of Pathology, Division of Diagnostics and Intervention, Oslo University Hospital, Oslo, Norway); Øystein Fodstad (Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway; Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway); Øystein Garred (Department of Pathology, Oslo University Hospital, Oslo, Norway); Gry Aarum Geitvik (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway); Gunhild Mari Mælandsmo (Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway; Department of Pharmacy, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway); Hege G. Russnes (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway; Department of Pathology, Oslo University Hospital, Oslo, Norway); Therese Sørlie (Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway); Ole Christian Lingjærde (Department of Computer Science, University of Oslo, Oslo, Norway); Helle Kristine Skjerven (Breast and Endocrine Surgery, Department of Breast and Endocrine Surgery, Vestre Viken Hospital, Drammen, Norway); Britt Fritzman (Østfold Hospital, Østfold, Norway).

In Silico Annotation of Candidate Causal Variants

Guide to results table (Supplementary Table 5) and UCSC Genome Browser session
Each candidate causal SNP has been annotated with publicly available genomic data in order to highlight potentially functional variants, prioritise experimental validation, and predict target genes. Annotations fall into categories relating to putative effects on transcription factors, regulatory element activities, expression quantitative trait loci (eQTL) and target gene prediction. For each variant, a link to the UCSC Genome Browser is provided that shows a 1 Mb region with relevant genomic data.

Transcription factors

Information regarding potential effects on transcription factor recognition sequences was obtained from the ENCODE-Motifs resource (http://compbio.mit.edu/encode-motifs)[1] using VCFtools v0.1.11 to access the downloaded HaploReg v4.0 database[2]. The impact each variant has on the position weight matrix for specific transcription factors is expressed as '+' or '-' for strengthened or weakened motifs relative to elements carrying the reference allele, respectively. Processed transcription factor ChIP-seq peak data for breast cell types were downloaded from ENCODE and other publications via NCBI GEO, in BED or NarrowPeak format, converted to the hg19 assembly using LiftOver if required, and given a standardised naming system (format = "*celltype;target*" in Supplementary Table 5). More details about the overlapping binding sites may be found within browser session track "TF-chip peaks overlapping candidate SNP" where TF-ChIP-seq peaks are named in the format "*Biosample_term_name, Experiment_target, Biosample_treatments, Biological_replicate(s), File_accession*". All ChIP-seq datasets are listed in Supplementary Table 6. Variants were assessed for overlap with ChIP-seq peaks using BedTools v2.25.0[3].

Regulatory features

Histone signatures derived from histone modification ChIP-seq experiments on breast cell types carried out by ENCODE, NIH Roadmap Epigenomics, and other published studies were obtained and formatted as for ChIP-seq data. Histone modification peaks overlapping candidate causal variants are represented as "*celltype;histone_mark*" in Supplementary Table 5 and "*Biosample_term_name, Experiment_target, Biosample_treatments, Biological_replicate(s), File_accession*" in the browser track "Histone modification ChIP-seq peaks overlapping candidate SNP". BedTools was used to intersect variants with histone signatures including commonly used marks associated with enhancers (H3K4me1, H3K4me2 and H3K27ac) and promoters (H3K4me3 and H3K9ac). Chromatin Hidden Markov Modelling (ChromHMM) states were obtained for breast cells from Roadmap (HMEC and myoepithelial cells) and published MCF7 data[4] and filtered for states corresponding to 'enhancers' (Roadmap 25-state E13, E14, E15, E16, E17, E18) or 'promoters' (Roadmap 25-state E1, E2, E3, E4, E22, E23). Chromatin state features containing candidate variants are represented as "*celltype;chromatin_state*". Chromatin accessibility data obtained from ENCODE, Roadmap and other published sources via NCBI GEO measured using DNase-seq and FAIRE-seq for relevant breast cell types were also tested for overlap with candidate causal variants. Intersected regions are reported in the format "*celltype;method*". Scores based on RegulomeDB[5] are presented for available SNPs (based on dbSNP141), where lower scores are increasingly likely to be functional (http://regulomedb.org/help#score).

eQTL

Genes showing expression levels correlated with query SNPs are shown in the column headed 'eQTL_target_all'. All genes reported to be associated with genotype in studies including GTEx version 6[6] (expression is any GTEx tissue, *P* value cutoff 10^{-6}) and Westra *et al*.,[7] (expression in whole blood). Genes associated with genotype in GTEx breast samples (N=186) are listed in the column "eQTL_GTEx.breast". eQTL data from TCGA and METABRIC studies for relevant variants are also presented (format "*Associated_gene:EffectAllele:EffectDirection:P_value*").

Other genomic features

Chromosomal position, the lead variant for the associated locus, and potentially conflicting rsIDs (assessed as overlapping at the query position) are given for each variant. GWAS tagSNPs were downloaded from the UCSC Table Browser (December 2015) and associated traits are listed if the tagSNP is within a 10 kb window of the candidate causal variant. NCBI RefSeq gene annotations were downloaded from the UCSC Table Browser and BedTools was used to determine overlapping genes ("Overlapping_RefGene"). The nearest RefGene transcription start site is also presented, given in the format "RefGeneTSS|distance". Basic genomic annotations such as intergenic, intronic, exonic, and untranslated regions based on RefSeq gene annotations were determined for each variant.

Target Gene Prediction

The column headed "Predicted_target_gene" lists genes predicted by various methods to be targets of, or the expression of which is associated with, regulatory elements in which the candidate causal variant lies. The reported gene is listed with cell type and method in the format "target:cell:method". A database was created comprising publicly available data based on various methods aiming to link enhancers with target genes (Annex to Supplementary Table 5). Laboratory based experimental approaches include genome-wide Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET)[8], Hi-C[9], and other Chromosome conformation capture (3C)-based techniques. Computational resources designed to predict target promoters by correlation of gene expression with ChIP-seq signals at specific regulatory elements including IM-PET[10], PreSTIGE[11] and data from Hnisz *et al.*[12] are also included. These methods associate enhancers defined by histone modification ChIP-seq for H3K4me1 (PreSTIGE), H3K27ac (Hnisz), H3K4me1, H3K4me3 and H3K27ac (IM-PET) with gene expression signals measured by RNA-seq. FANTOM5[13] data representing enhancer-promoter cap analysis of gene expression (CAGE) expression correlation from all cell types were downloaded from http://enhancer.binf.ku.dk/. Target genes have been predicted for multiple cell types and all data were included in the database, and filtered for breast derived cell types for this analysis (see Key to Supplementary Table 5).

The following strategy was used to assign potential target genes to regulatory elements. The published computational methods (Hnisz, PreSTIGE and IM-PET) included target gene annotation in the reported data. For ChIA-PET and Hi-C data, interaction peaks were mapped to promoters defined as -1.0 kb to +0.1 kb around GENCODE (v19) transcription start sites. Enhancer definitions were used as reported for computational methods while for ChIA-PET and Hi-C were interpreted as any region interacting with a promoter (regardless of other enhancer annotation information such as histone modification or open chromatin). FANTOM5 target promoters were predefined and

tissue specificity was determined by intersecting “TSS associated enhancers” with tissue-specific sets of enhancers.

All data were formatted to enable intersection of test variants with “enhancers” as defined by each method using the Galaxy “intersect” tool[14]. Each enhancer-promoter assignment or interaction was represented as a single record along with details about potential target promoter, cell type, method, scoring and confidence statistics from the original publication. A set of query SNPs (or any loci with genomic positional information in BED format) could be queried into a custom Galaxy workflow leading to generation of a table of predicted gene targets and a link to the UCSC Genome Browser for visualisation.

UCSC Genome Browser session

A custom session has been uploaded to UCSC Genome Browser[15] to facilitate exploration of breast cancer risk associated variation and implicated regulatory features. This can be accessed via the hyperlink (ie. “browser”) in the functional annotation *xlsx* file. All standard Genome Browser data and functions are then available, including track display options (eg. right click on a particular track to activate visualisation options), highlighting regions (shift and mouse over region of interest), and the table browser (eg. to intersect or export data).

Within the session, Oncoarray candidate causal variants are shown in red, and names can be shown by activating “pack” mode (as for all tracks). Target gene prediction data from Hnisz, PreSTIGE and IM-PET shows enhancers depicted as black bars. The segment name revealed in ‘pack’ mode lists predicted target gene and cell-type (eg “WNT7B.MCF7”). ChIA-PET interactions, represented in BED12 format, have been filtered to remove duplicates and *trans*-chromosomal interactions. The interactions are shaded to reflect statistical confidence based on enrichment in the original experiment. ChIA-PET interaction names show the genomic co-ordinates of either-end of the interaction, the cell type (restricted to MCF7 for this analysis), the immunoprecipitation target, and the experimental replicate. Depicted interactions are restricted to those for which a candidate variant lies within an interaction “end” with the opposite end overlapping a TSS. All other interactions may be visualised by activating the standard ENCODE ChIA-PET track (\Regulation\ENCODE Chromatin Interactions Tracks\ChIA-PET from ENCODE/GIS-Ruan).

Chromatin interactions based on *in situ* Hi-C data from HMEC cells were downloaded from NCBI GEO (accession GSE63525)[9]. Annotated loops (representing potential enhancer-promoter interactions) processed by HiCCUPS were reformatted as BED files and tested for overlap with RefSeq promoters to assign potential target genes. Opposing ends of TSS-overlap loops were then annotated as ‘potential enhancers’. Specific loop regions which overlap BC risk candidate causal variants are depicted as black segments and named “*TSS_target.Celltype*”.

Various classes of genomic data representing regulatory features which harbour candidate variants are shown as separate tracks:

- *Histone modification ChIP-seq peaks overlapping candidate SNP*
- *DNase HS and FAIRE-seq peaks overlapping candidate SNP*
- *TF-chip peaks overlapping candidate SNP*

As mentioned above, changing the track display to 'pack' mode will show details of the overlapping peak in the format:
"*Biosample_term_name, Experiment_target, Biosample_treatments, Biological_replicate(s), File_accession*".

A representation of all TF and histone ChIP-seq, DNase-seq, and FAIRE-seq data tested for overlap with candidate variants is shown in three histogram tracks (computed with BedTools genomeCoverageBed). These show the summed peak density at each genomic position and allow simple visualisation of loci with relative abundance of regulatory features.

Tracks for Roadmap Epigenomics Chromatin state models (based on imputed data - 25 state, 12 marks, 127 epigenomes) were generated for breast Myoepithelial and HMEC cells. Chromatin states were separated and colour coded for states related to enhancers, promoters, and transcribed regions.

The bottom track ("Oncoarray SNPs") shows all directly genotyped and imputed SNPs passing quality control (imputation r^2>0.3) as black ticks. SNPs from dbSNP build 138 with a MAF > 0.01 in European samples which were not informative are shown in red.

Funding

Genotyping of the OncoArray was funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research (GPH-129344), the *Ministère de l'Économie, de la Science et de l'Innovation du Québec* through Genome Québec, the Quebec Breast Cancer Foundation for the PERSPECTIVE project, the US National Institutes of Health (NIH) [1 U19 CA 148065 for the Discovery, Biology and Risk of Inherited Variants in Breast Cancer (DRIVE) project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN268201200008I], Cancer Research UK [C1287/A16563], Odense University Hospital Research Foundation (Denmark), the National R&D Program for Cancer Control - Ministry of Health & Welfare (Republic of Korea) [1420190], the Italian Association for Cancer Research [AIRC, IG16933], the Breast Cancer Research Foundation, the National Health and Medical Research Council (Australia) and German Cancer Aid [110837].
Genotyping of the iCOGS array was funded by the European Union [HEALTH-F2-2009-223175], Cancer Research UK [C1287/A10710, C1287/A10118, C12292/A11174], NIH grants (CA128978, CA116167, CA176785) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), an NCI Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201) the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, the *Ministère de l'Économie, Innovation et Exportation du Québec* (#PSR-SIIRI-701), Komen Foundation for the Cure, the Breast Cancer Research Foundation and the Ovarian Cancer Research Fund.

Combining the GWAS data was supported in part by NIH Cancer Post-Cancer GWAS initiative [1 U19 CA 148065] (DRIVE, part of the GAME-ON initiative). LD score regression analysis was supported by CA194393.

BCAC is funded by Cancer Research UK [C1287/A16563] and by the European Union via its Seventh Framework Programme [HEALTH-F2-2009-223175, (COGS)] and Horizon 2020 Research & Innovation Programme [633784 (B-CAST); 634935 (BRIDGES)]. CIMBA is funded by Cancer Research UK [C12292/A20861 and C12292/A11174].
This work was also funded by US National Cancer Institute (NCI) [CA27469, CA37517, CA54281, CA58860, CA63464, CA92044, CA97396, CA098758, CA116167, CA116201, CA125183, CA128931, CA132839, CA140286, CA164973, CA173531, CA176785, CA177150, CA192393, P01-CA87969, P20-GM103534, P30-CA014089, P30-CA016056, P30-CA023108, P30-CA051008, P30-CA68485, P30-CA168524, P50-CA136393, R01-CA77398, R01-CA100374, R01-CA121941, R01-CA122443, R01-CA140323, R01-CA142996, 1R01-CA149429, 1R-03CA130065, RC4-CA153828, UM1-CA164917, UM1-CA164920, UM1-CA176726, UM1-CA186107, 1U01-CA161032, 5U01-CA113916, U01-CA199277, U10-CA180822, U10-CA180868, U19-CA148065, U19-CA148127, 5-U19-CA148112, Intramural Research Programs], support services contracts with Westat, Inc, Rockville, MD [NO2-CP-11019-50, N02-CP-65504], US National Institute of Environmental Health Sciences (NIEHS) [Z01ES044005, Z01ES102245, Z01ES049033, Intramural Research Program], US National Institutes of Health (NIH) [P41-GM103504, U41-HG006623], Research Council of Lithuania [SEN-18/2015], Breast Cancer Research Foundation (USA), Cancer Association of South Africa (CANSA), Morris and Horowitz Familes Endowed Professorship, Asociacion Española Contra el Cancer [AECC08], Red Temática de Investigación Cooperativa en Cáncer [RD06/0020, RD06/0020/1051, RD06/0020/1060, RD12/0036/0006, RD12/0036/008], Fondo de Investigaciones Sanitarias (FIS) [PI05/2275, PI08/1120, PI10/00748, PI10/01422, PI11/00923, PI12/00070, PI12/02125, PI13/00061, PI13/00285, PI13/01136, PI13/01162, PI1500854], ISCIII (Spain) [PIE13/00022-ONCOPROFILE, JR14/00017, 15/00059], Fundación Mutua Madrileña (FMMA) [SAF2010-20493], Catalan Health Institute and Autonomous Government of Catalonia [2009SGR290, 2014SGR364] Office of the Director - US National Institutes of Health, Italian Association of Cancer Research (AIRC) [IG 2012 Id.12821; IG 2014 Id.15547], Ministero della Salute (Italy), Italian citizens' "5x1000" funding (to the Istituto Oncologico Veneto, Fondazione IRCCS Istituto Nazionale Tumori, IRCCS AOU San Martino - IST), European Social Fund (ESF), Greek National Strategic Reference Framework (NSRF) [SYN11_10_19 NBCA], German Cancer Research Centre (DKFZ), Cancer Research UK [C490/A10124, C570/A16491, C1275/A11699, C1275/A15956, C1275/A19187, C1275/C22524, C1287/A10118, C1287/A11990, C8221/A19170, 14136, C8197/A16565], UK National Institute for Health Research (NIHR), University of Kansas Cancer Center, Kansas Bioscience Authority Eminent Scholar Program, Fisher Center for Hereditary Cancer and Clinical Genomics Research, Swing Fore the Cure, Research Fund - Flanders (FWO), European Regional Development (FEDER), Helsinki University Hospital Research Fund, Academy of Finland [266528], Finnish Cancer Society, Sigrid Juselius Foundation, Dutch Cancer Society [NKI-1998-1854, NKI-2004-3088, NKI-2007-3756, RUL-1997-1505, NKI-2007-3839, NKI-2009 4363, DDHK-2004-3124, DDHK-2009-4318], Netherlands Organization of Scientific Research [NWO 91109024], Pink Ribbon [110005], BBMRI [NWO 184.021.007/CP46], Hungarian Research Grants [KTIA-OTKA CK-80745, OTKA K-112228], German Cancer Aid (70492, 70-2892-BR I, 106332, 108253, 108419, 110826, 110828), Liga Portuguesa Contra o Cancro, Cancer Australia, Polish Committee of Scientific Research

(PBZ_KBN_122/P05/2004), National Breast Cancer Foundation (Australia), National Health and Medical Research Council (Australia) [199600, 209057, 251553,400281, 400413, 504711, 572530, 1041801], Queensland Cancer Fund, Cancer Council New South Wales, Cancer Council Victoria, Cancer Council Tasmania, Cancer Council South Australia, Cancer Council Western Australia, Cancer Institute NSW, VicHealth (Australia), Victorian Cancer Agency (Australia), David F. and Margaret T. Grohne Family Foundation, Jewish General Hospital Weekend to End Breast Cancer, Quebec Ministry of Economic Development, Innovation and Export Trade, Ministry of Health of the Czech Republic [MMCI, 00209805], European Regional Development Fund, State Budget of the Czech Republic (RECAMO) [CZ.1.05/2.1.00/03.0101], Charles University in Prague project [UNCE204024], Robert and Kate Niehaus Clinical Cancer Genetics Initiative, Andrew Sabin Research Fund, Russian Federation for Basic Research [14-04-93959, 15-04-01744], Ohio State University Comprehensive Cancer Center, Istituto Toscano Tumori (ITT) [2011-2013], Israel Cancer Association, Israeli Inherited Breast Cancer Consortium, Swedish Cancer Society, Swedish Cancer Foundation, Ralph and Marion Falk Medical Research Trust (Sweden), Entertainment Industry Fund National Women's Cancer Research Alliance, UCSF Cancer Risk Program, Helen Diller Family Comprehensive Cancer Center, Susan G. Komen Foundation for the Cure (FAS0703856), Basser Research Center for BRCA, Frieda G. and Saul F. Shapira BRCA-Associated Cancer Research Program, Hackers for Hope Pittsburgh, Ovarian Cancer Research Fund (OCRF) [258807, POE/USC/01.12], American Cancer Society [SIOP-06-258-01-COUN], US National Center for Advancing Translational Sciences (NCATS) [UL1TR000124], ELAN-Fond (University Hospital of Erlangen, Germany), Cancer Research UK, UK National Health Service (NHS), Breast Cancer Now (UK), The Institute of Cancer Research (UK), NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre, Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust in partnership with King's College London, Oxford Biomedical Research Centre, Xunta de Galicia (Spain) [10CSA012E], Ministerio de Economía y Competitividad (Spain) [Feder-Innterconnecta-ITC-20133101], Ministerio de Sanidad, Servicios Sociales e Igualdad (Spain) [EC11-192], Dietmar-Hopp Foundation, Helmholtz Society, Canadian Institutes of Health Research, University of Crete, Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), Agence Nationale de la Recherche (ANR). Chief Physician Johan Boserup and Lise Boserup Fund (Denmark), Danish Medical Research Council, Herlev and Gentofte Hospital (Copenhagen), American Cancer Society, California Breast Cancer Act of 1993, California Breast Cancer Research Fund [97-10500], California Department of Public Health, Lon V Smith Foundation [LVS39420], Against Breast Cancer Registered Charity No. 1121258 (UK), European Commission [DG-SANCO], European Union Marie Skłodowska-Curie Individual Fellowship [MSCA-IF-2014-EF-656144], International Agency for Research on Cancer (IARC), Institut Gustave Roussy (France), Mutuelle Générale de l'Education Nationale (France), Institut National de la Santé et de la Recherche Médicale (INSERM) (France), Hellenic Health Foundation (Greece), Stavros Niarchos Foundation (Greece), National Research Council (Italy), Dutch Ministry of Public Health Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds (Netherlands), Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (Netherlands), Regional Governments of Andalucía, Asturias, Basque

Country, Murcia and Navarra (Spain), Medical Research Council (UK) [1000143, MR/M012190/1], Baden Württemberg Ministry of Science Research and Arts (Germany), Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London, Claudia von Schilling Foundation for Breast Cancer Research (Germany), Lower Saxonian Cancer Society (Germany), Nordic Cancer Union, Friends of Hannover Medical School (Germany), Rudolf Bartling Foundation (Germany), Russian Foundation for Basic Research [14-04-97088], Stockholm County Council, Karolinska Institutet, Gustav V Jubilee Foundation (Sweden), Bert von Kantzows Foundation (Sweden), Märit and Hans Rausings Initiative Against Breast Cancer (Sweden), special Government Funding (EVO, Finland) [for research at Oulu Univeristy and Kuopio University Hospitals], Kuopio University Hospital (Finland), Cancer Fund of North Sav (Finland), Finnish Cancer Organizations, University of Eastern Finland, US Army Medical Research and Materiel Command [DAMD17-01-1-0729], Stichting tegen Kanker (Belgium), Research Centre for Genetic Engineering and Biotechnology "Georgi D. Efremov" (Macedonia), German Academic Exchange Program (DAAD) Hamburg Cancer Society, German Federal Ministry of Education and Research (BMBF) [01KH0402, RUS08/017, 01KW9975/5, 01KW9976/8, 01KW9977/0 and 01KW0114], Robert Bosch Foundation, Stuttgart, Germany, Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bochum, Germany, Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany, David F. and Margaret T. Grohne Family Foundation (USA), European Research Council [ERC-2011-294576], Berta Kamprad Foundation, Gunnar Nilsson, Research Council of Norway [193387/V50, 193387/H10], South Eastern Norway Health Authority [39346, 27208], Norwegian Cancer Society [419616-71248-PR-2006-0282], K.G. Jebsen Centre for Breast Cancer Research [2012-2015], Finnish Cancer Foundation, Academy of Finland [250083, 122715, 251314], Finnish Cancer Foundation, Sigrid Juselius Foundation, University of Oulu, University of Oulu Support Foundation, Biobanking and Biomolecular Resources Research Infrastructure (Netherlands) [BBMRI-NL CP16], Breast Cancer Campaign (UK) [2010PR62, 2013PR044], Agency for Science, Technology and Research of Singapore (A*STAR), Susan G. Komen Breast Cancer Foundation (USA), Sheffield Experimental Cancer Medicine Centre (UK), European Regional Development Fund and Free State of Saxony (Germany) [LIFE - Leipzig Research Centre for Civilization Diseases: 713-241202, 713-241202, 14505/2470, 14575/2470].

Acknowledgements

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. We specifically thank Andrew Berchuck, Rosalind A. Eeles, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch, Sylvie LaBoissière, Frederic Robidoux, Borge G. Nordestgaard, Sharon A. Windebank, Christopher A. Hilker, Blood bank Sanquin (The Netherlands), ABCTB Investigators (Christine Clarke, Rosemary Balleine, Robert Baxter,Stephen Braye, Jane Carpenter, Jane Dahlstrom, John Forbes, Soon Lee, Debbie Marsh, Adrienne Morey, Nirmala Pathmanathan, Rodney Scott, Allan Spigelman, Nicholas Wilcken and Desmond Yip), Eileen Williams, Elaine Ryder-Mills, Kara Sargus, Allyson Thomson, Christobel Saunders, Terry Slevin, BreastScreen

Western Australia, Elizabeth Wylie, K. Landsman, N. Gronich, A. Flugelman, W. Saliba, E. Liani, I. Cohen, S. Kalet, V. Friedman, O. Barnet, Niall McInerney, Gabrielle Colleran, Andrew Rowan, Angela Jones, Victor Muñoz Garzón, Alejandro Novo Domínguez, Sara Miranda Ponte, Carmen M Redondo, Maite Peña Fernández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza, Francisco Gude Sampedro, José Antúnez, Máximo Fraga and the staff of the Department of Pathology and Biobank (University Hospital Complex of Santiago-CHUS, Instituto de Investigación Sanitaria de Santiago, IDIS, Xerencia de Xestion Integrada de Santiago-SERGAS, Spain), Joaquín González-Carreró and the staff of the Department of Pathology and Biobank (University Hospital Complex of Vigo, Instituto de Investigacion Biomedica [IBI] Orense-Pontevedra-Vigo, SERGAS, Vigo, Spain), Peter Bugert, Angela Brooks-Wilson, Agnes Lai and Anne Grundy. Styliani Apostolaki, Anna Margiolaki, Georgios Nintos, Maria Perraki, Georgia Saloustrou,Georgia Sevastaki, Konstantinos Pompodakis, Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, Dorthe Kjeldgård Hansen, Danish Cancer Biobank, Guillermo Pita, Nuria Álvarez, Alicia Barroso, Rosario Alonso, Pilar Zamora, the Human Genotyping-CEGEN Unit (CNIO), CPS-II Investigators, CPS-II Study Management Group, cancer registries supported by the Centers for Disease Control and Prevention National Program of Cancer Registries and National Cancer Institute Surveillance Epidemiology and End Results program, the CTS Steering Committee (Leslie Bernstein, Susan Neuhausen, James Lacey, Sophia Wang, Huiyan Ma, Jessica Clague DeHart, Dennis Deapen, Rich Pinder, Eunjung Lee, Pam Horn-Ross, Peggy Reynolds, Christina Clarke Dur, David Nelson, Hoda Anton-Culver, Argyrios Ziogas,Hannah Park, Fred Schumacher), EPIC investigators, Hartwig Ziegler, Sonja Wolf, Volker Hermann, Katja Butterbach, Stefanie Engert, Heide Hellebrand, Sandra Kröber, LIFE - Leipzig Research Centre for Civilization Diseases (Markus Loeffler, Joachim Thiery, Matthias Nüchter, Ronny Baber), The GENICA Network (Yon-Dschun Ko, Christina Justenhoven, Christian Baisch, Beate Pesch, Sylvia Rabstein, Anne Lotz, Volker Harth), Ursula Eilber, Kelly Kohut, Patricia Gorman, Maria Troy, Michael Bremer, Rainer Fagerholm, Kirsimari Aaltonen, Karl von Smitten, Irja Erkkilä, Peter Hillemanns, Hans Christiansen, Johann H. Karstens, Shamil Gantsev, The Swedish Medical Research Counsel, Eija Myöhänen, Helena Kemiläinen, Gilian Peuteman, Thomas Van Brussel, EvyVanderheyden, Kathleen Corthouts, Milena Jakimovska, Katerina Kubelka, Mitko Karadjozov, Andrej Arsovski, Liljana Stojanovska, Petra Seibold, Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers, Stefan Nickels, Siranoush Manoukian, Lidia Pezzani, Irene Feroce, Martine Tranchant, Marie-France Valois, Annie Turgeon, Lea Heguy, NBCS collaborators (Kristine K.Sahlberg, Lars Ottestad, Rolf Kåresen, Anita Langerød, Ellen Schlichting, Marit Muri Holmen, Toril Sauer, Vilde Haakensen, Olav Engebråten, Bjørn Naume, Cecile E. Kiserud, Kristin V. Reinertsen, Åslaug Helland, Margit Riis, Ida Bukholm, Per Eystein Lønning, Oslo Breast Cancer Research Consortium), the following US state cancer registries (AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY), Arja Jukkola-Vuorinen, Mervi Grip, Saila Kauppila, Meeri Otsukka, Leena Keskitalo, Kari Mononen, Teresa Selander, Nayana Weerasooriya, E. Krol-Warmerdam, J. Blom, J. Molenaar, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner Sonja Oeser, Silke Landrith, Petra Bos, Jannet Blom, Ellen Crepin, Elisabeth Huijskens, Anja Kromwijk-Nieuwlaat, Annette Heemskerk, Sue Higham, Helen Cramp, Dan Connley, Ian Brock, Sabapathy Balasubramanian, Malcolm W.R. Reed, the SEARCH and EPIC teams, UK National Institute for Health Research Biomedical Research Centre, the SUCCESS Study teams in Munich, Duessldorf, Erlangen and Ulm

(Germany), Ewa Putresza, Irene Masunaka, the WHI investigators, Maggie Angelakos, Judi Maskiell, Gillian Dite, Helen Tsimiklis, Vilius Rudaitis, Laimonas Griškevičius, Janis Eglitis, Anna Krilova, Aivars Stengrevics, Yuan Chun Ding, Linda Steele, Irene Feroce, Daniela Furlan, Viviana Gismondi Riccardo Dolcetti, Gabriele Capone, Anna Laura Putignano, Giuseppe Guannini, Antonella Savarese, Aline Martayan, Stefania Tommasi, Brunella Pilato, the Cogentech Cancer Genetic Test Laboratory (Italy), Riccardo Dolcetti, Barbara Pasini, Francesca Vignolo-Lutati, Sue Healey, JoEllen Weaver, Betsy Bove, Marta Santamariña, Ana Blanco, Miguel Aguado, Uxía Esperón, Belinda Rodríguez, GEMO Collaborating Centers (Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Hospices Civils de Lyon & Centre Léon Bérard ; Service de Génétique Oncologique, Institut Curie, Paris; Service de Génétique & BB-0033-00074, Gustave Roussy, Villejuif; Centre Léon Bérard, Lyon; Centre François Baclesse, Caen; Institut Paoli Calmettes, Marseille; CHU Arnaud-de-Villeneuve, Montpellier; Centre Oscar Lambret, Lille; Centre Paul Strauss, Strasbourg; . Institut Bergonié, Bordeaux; Institut Claudius Regaud, Toulouse; CHU Grenoble; CHU Dijon; CHU St-Etienne; Hôtel Dieu Centre Hospitalier, Chambéry; Centre Antoine Lacassagne, Nice; CHU Limoges; CHU Nantes; CHU Bretonneau, Tours; Groupe Hospitalier Pitié-Salpétrière, Paris; CHU Vandoeuvre-les-Nancy; CHU Besançon; Creighton University, Omaha, USA), GEMO Investigators (Olga Sinilnikova†, Carole Verny-Pierre, Mélanie Léone, Nadia Boutry-Kryza, Alain Calender, Sophie Giraud, Bruno Buecher, Claude Houdayer, Etienne Rouleau, Lisa Golmard, Agnès Collet, Virginie Moncoutier, Muriel Belotti, Antoine de Pauw, Camille Elan, Catherine Nogues, Emmanuelle Fourme, Anne-Marie Birot, Marine Guillaud-Bataille. Centre Jean Perrin, Clermont–Ferrand, Nancy Uhrhammer, Valérie Bonadona, Sandrine Handallou, Agnès Hardouin, Pascaline Berthet, Dominique Vaur, Hagay Sobol, Violaine Bourdon, Tetsuro Noguchi, Audrey Remenieras, François Eisinger, Pascal Pujol, Jean-Philippe Peyrat, Joëlle Fournier, Françoise Révillion, Philippe Vennin†, Claude Adenis, Danièle Muller, Jean-Pierre Fricker, Emmanuelle Barouk-Simonet, Françoise Bonnet, Virginie Bubien, Nicolas Sevenet, Michel Longy, Christine Toulas, Rosine Guimbaud, Laurence Gladieff, Viviane Feillel, Dominique Leroux, Hélène Dreyfus, Christine Rebischung, Magalie Peysselon, Fanny Coron, Fabienne Prieur, Marine Lebrun, Caroline Kientz, Sandra Fert Ferrer, Marc Frénay, Laurence Vénat-Bouvet, Capucine Delnatte, Isabelle Mortemousque, Florence Coulet, Chrystelle Colas, Florent Soubrier, Mathilde Warcoin, Johanna Sokolowska, Myriam Bronner, Marie-Agnès Collonge-Rame, Alexandre Damette, Henry T. Lynch, Carrie L. Snyder, Ilse Coene, Brecht Crombez, Alicia Tosar, Paula Diaque, Kristiina Aittomäki, Taru A. Muranen, Kirsimari Aaltonen, Irja Erkkilä, Virpi Palola, Comprehensive Cancer Centre Netherlands, Comprehensive Centre South, Netherlands Cancer Registry, Dutch Pathology Registry, HEBON Collaborating Centers (Netherlands Cancer Institute, Erasmus Medical Center - Rotterdam, Leiden University Medical Center, Radboud University Nijmegen Medical Center, University Medical Center Utrecht, Amsterdam Medical Center, VU University Medical Center - Amsterdam, University Hospital Maastricht, The Netherlands Foundation for the detection of hereditary tumours, The Netherlands Comprehensive Cancer Organization [IKNL], the Dutch Pathology Registry [PALGA], Groningen University Medical Center - Groningen), HEBON Investigators (F.E. van Leeuwen, S. Verhoef, N.S. Russell, J.L. de Lange, R. Wijnands, J.M. Collée, C.H.M. van Deurzen, I.M. Obdeijn, T.C.T.E.F. van Cronenburg, A.R. Mensenkamp, R.B. van der Luijt, C.C. van der Pol, J.J.P. Gille, Q. Waisfisz, H.F. Vasen, S. Siesling, J.Verloop, L.I.H. Overbeek. E.B. Gómez-Garcia, A.H. van der Hout, M.J. Mourits, G.H. de Bock) Janos Papp, Tibor Vaszko, Aniko Bozsik, Judit Franko, Maria Balogh, Gabriella Domokos, Judit Ferenczi, the Oncogenetics

Group (VHIO) and the High Risk and Cancer Prevention Unit (University Hospital Vall d'Hebron, the Cellex Foundation, Spain), Gabriel Capella, Martine Dumont, Catarina Santos, Pedro Pinto, Heather Thorne, Eveline Niedermayr, the kConFab Clinical Follow Up Study, Csilla Szabo, Lenka Foretova, Eva Machackova, Michal Zikan, Petr Pohlreich, Zdenek Kleibl, Anne Lincoln, Lauren Jacobs, Australia New Zealand NRG Oncology Group investigtors, the Ontario Cancer Genetics Network, Leigha Senter, Kevin Sweet, Caroline Craven, Julia Cooper, Michelle O'Conor, the Meirav Comprehensive Breast Cancer Center team (Sheba Medical Center), Åke Borg, Helena Jernström, Karin Henriksson, Katja Harbst, Maria Soller, Ulf Kristoffersson, Anna Öfverholm, Margareta Nordling, Per Karlsson, Zakaria Einbeigi, Anna von Wachenfeldt, Annelie Liljegren, Brita Arver, Gisela Barbany Bustinza, Johanna Rantala, Beatrice Melin, Christina Edwinsdotter Ardnor, Monica Emanuelsson, Maritta Hellström Pigg, Richard Rosenquist, Marie Stenmark-Askmalm, Sigrun Liedgren Cecilia Zvocec, Qun Niu, Robert Nussbaum, Beth Crawford, Kate Loranger, Julie Mak, Nicola Stewart, Robin Lee, Amie Blanco, Peggy Conrad, Salina Chan, Simon Gayther, Carole Pye, Patricia Harrington, Eva Wozniak, Geoffrey Lindeman, Marion Harris and Martin Delatycki (Victorian Familial Cancer Trials Group), Sarah Sawyer, Rebecca Driessen and Ella Thompson. MCCS cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. The authors assume full responsibility for analyses and interpretation of these data. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the BCFR. We pay tribute to Olga M. Sinilnikova, who with Dominique Stoppa-Lyonnet initiated and coordinated GEMO until she sadly passed away on the 30th June 2014.

OncoArray genotype calling and quality control

Of the 568,712 variants selected for genotyping on OncoArray, 533,631 were successfully manufactured on the array (including 778 duplicate probes). OncoArray genotyping of BCAC and CIMBA samples was conducted at six sites. Details of the genotyping calling for the OncoArray are described in more detail elsewhere[16]. Briefly, we developed a single calling pipeline that was applied to more than 500,000 samples. An initial cluster file was generated using from 56,284 samples, selected over all the major genotyping centres and ethnicities, using the Gentrain2 algorithm. Variants likely to have problematic clusters were selected for manual inspection using the following criteria: call rate below 99%, minor allele frequency (MAF) <0.001, poor Illumina intensity and clustering metrics, deviation from the MAF observed in the 1000 Genomes Project using the criterion: $\frac{(|p_1-p_0|-0.01)^2}{((p_1+p_0)(2-p_1-p_0))} > C$, where p_0 and p_1 are the minor frequencies in the 1000 Genome Project and Oncoarray datasets, respectively, and C=0.008. (This latter criterion is approximately equivalent to excluding SNPs on the basis of a Chi-square statistic of 16 for the difference in allele frequencies, assuming 1,000 samples in each group). This resulted in manual adjustment of the cluster file for 3,964 variants, and the exclusion of 16,526 variants. The final cluster file was then applied to the full dataset.

We excluded SNPs with a call rate <95% in any consortium, not in Hardy-Weinberg equilibrium ($P<10^{-7}$ in controls, or $P<10^{-12}$ in cases) or with concordance <98% among 5,280 duplicate pairs. For the imputation, we additionally excluded SNPs with a MAF<1% and a call rate <98% in any consortium, SNPs that could not be linked to the 1000 Genomes Project reference, those with MAF for Europeans that differed from that for the 1000 Genomes Project and a further 1,128 SNPs where the cluster plot was judged to be not ideal. Of the 533,631 SNPs which were manufactured on the array, 494,763 passed the initial QC and 469,364 were used in the imputation (see below).

For BCAC, we excluded probable duplicate samples and close relatives within each study, and probable duplicates between studies. These were identified by identity by state (IBS) analysis using a set of approximately 38,000 uncorrelated ($r^2<0.1$) SNPs for OncoArray and iCOGS and 16,000 SNPs for GWAS. Based on inspection of the distribution of IBS values, we identified first-degree relative pairs using the criterion 0.82<IBS<0.90 for OncoArray and 0.85<IBS<0.90 for iCOGS; similar criteria were used for each GWAS (with limits depending on the IBS distribution in that study).

We applied LD score regression to the summary results from GWAS, iCOGS and OncoArray to assess the evidence of overlap in individuals between the three datasets. We conducted three pair-wise cross-trait regression analyses (GWAS-iCOGS, GWAS-OncoArray and iCOGS-OncoArray) and used the intercept from the regression analysis to estimate the amount of overlap[17]. Assuming that the phenotypic correlation is 1 (that is, a case is a case in all datasets and a control is a control in all datasets), we found that for GWAS-iCOGS, the estimated overlap was 1.5% of individuals, for GWAS-OncoArray, the estimated overlap was 3.8% of individuals, and for iCOGS-OncoArray, the estimated overlap was 0.2% of individuals. It is unlikely that this degree of overlap would have influenced the results obtained from our analyses.

We also excluded samples with a call rate <95% and samples with extreme heterozygosity (>4.9 standard deviations from the mean for the reported ethnicity). Ancestry analysis was performed using a standardized approach in which 2,318 ancestry informative markers with minor allele frequencies of 0.05 on a subset of ~66,000 samples including 505 Hapmap 2 samples. The contribution of each of the three major continental ancestry groups (European, Asian and African) was estimated by mapping each individual to regions of a triangle based on the first two principal components, as implemented in the software package FastPop (http://sourceforge.net/projects/fastpop/)[18]. Individuals were thus classified into 4 groups: European (defined as >80% European ancestry), East Asian (>40% Asian ancestry), African (>20% African ancestry) and other (not fulfilling any of the above criteria)[16]. Of the 152,492 samples genotyped, the final dataset consisted of 142,072 samples, of which 9,655 ER-negative cases and 45,494 controls of European origin had not been included in a previous GWAS and had not been genotyped using iCOGS and were included in this analysis.

For the CIMBA samples we excluded individuals of non-European ancestry using multi-dimensional scaling. For this purpose we selected 30,733 uncorrelated autosomal SNPs (pair-wise $r^2< 0.10$) to compute the genomic kinship between all pairs of *BRCA1* and *BRCA2* carriers, along with 267 HapMap samples (CHB, JPT,

YRI and CEU). These were converted to distances and subjected to multidimensional scaling. Using the first two components, we calculated the proportion of European ancestry for each individual and excluded samples with >27% non-European ancestry to ensure that samples of Ashkenazi Jewish ancestry were included in the final sample.

Global Genomic Enrichment Analyses (further details)

We created a "full baseline model" as previously described[19] that included 52 "baseline" genomic features (24 non-cell-type specific publicly available annotations, a 500-bp window around each of the 24 annotations and a 100-bp window around each of four ChIP-seq peaks) and one category containing all SNPs. We estimated the enrichment for these 53 functional categories in a single multivariable LD score regression analysis.

We subsequently performed analyses using cell-type specific annotations for the four histone marks H3K4me1, H3K4me3, H3K9ac and H3K27ac across 27-81 cell types, depending on histone mark, giving a total of 220 cell-type specific marks[19]. We estimated the enrichment for each of these marks after adjusting for the baseline annotations by running 220 LD score regressions, each adding a different histone mark to the baseline model. We observed no associations after adjusting for 220 tests

Pathway Enrichment Analyses (further details)

Pathway enrichment analysis was performed to identify pathways associated with ER-negative breast cancer risk, pointing to biological hypotheses that can be further tested experimentally.

The pathway gene set database used contains pathway gene sets from Reactome[20], NCI Pathway Interaction Database[21], GO (Gene Ontology) biological process[22], HumanCyc[23], MSigdb[24], NetPath[25] and Panther[26]. GO pathways inferred from electronic annotation terms were excluded. Some manual annotation was performed on the pathway gene set database where annotation errors from public data were discovered. In particular, in several pathways, the PDPK1 gene was mistakenly entered as PDK1 gene and was manually corrected. The same pathway (e.g. apoptosis) may be defined in two or more databases with potentially different sets of genes, and all versions of these duplicate/overlapping pathways were included.

Gene information (hg19) was downloaded from the ANNOVAR[27] website (http://www.openbioinformatics.org/annovar/). Some pathways include genes that are also grouped closely together in the genome and are thus are likely to share the significance of a single SNP, which would artificially increase the pathway significance in our analysis. This was the case for pathways including histone genes. Thus, we selected representative SNP-gene associations to control for this effect (chr6:26055031 for HIST1, chr1:120904839, 149864043 for HIST2, chr1: 228615251 for HIST3 and chr12: 14919727 for HIST4).

Although there are several methods for pathway enrichment analysis, we chose the GSEA approach as it is one of the most established methods that is threshold free;

many other methods such as SRT, ALIGATOR and Plink set-based test require an arbitrary p-value threshold to be defined for SNPs and applied before pathway analysis.

To focus on pathway enrichment analysis results about which we were most confident, we implemented a number of filters. First, only pathways with positive ES and containing at least one gene linked to a significant SNP ($P<5x10^{-8}$) were retained for subsequent analysis. Second, we defined an ES threshold (ES≥0.41) based on a comparison with a gold standard pathway enrichment analysis we previously performed on the iCOGS data alone and where we were able to analytically compute FDR values by shuffling case/control labels (this was not computationally feasible with the more complex meta-analysis scheme used in this paper).

We chose the true positive rate (TPR) threshold by varying the TPR in steps of 0.1 and observing how the FPR changed. A TPR of 0.1 resulted in a very low FPR (0.02), but we considered this to be unduly conservative as it resulted in a small number of pathways (37, clustered into 8 themes) and excluded many pathways known to be involved in breast cancer. A TPR of 0.20 (FPR = 0.14) gave a reasonable balance between the true and false positive rates, while including pathways known to be involved in breast cancer. Thus this threshold was chosen for this study. A TPR of 0.3 gave an FPR of 0.30, which we considered high; further, the resulting additional pathways included (in addition to those included at TPR=0.2) were weaker (i.e. they had worse enrichment scores [ES<0.41] and had relatively very few genes included) than pathways appearing at lower FPRs (and TPRs). We rejected TPR thresholds >0.3 because each gave an FPR that was larger than the TPR.

Finally, we performed an in depth literature search on all resulting pathways to confirm their relevance to breast cancer biology, applying the following criteria: 1) reported in at least one of five published breast cancer pathway analyses[28-32]; or 2) reported elsewhere in the literature to be involved in breast cancer. We also removed pathways that were significant due to incorrect gene function annotation.

References

1. Kheradpour, P. & Kellis, M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res* **42**, 2976-87 (2014).
2. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**, D930-4 (2012).
3. Quinlan, A.R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**, 11 12 1-34 (2014).
4. Taberlay, P.C., Statham, A.L., Kelly, T.K., Clark, S.J. & Jones, P.A. Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. *Genome Res* **24**, 1421-32 (2014).

5. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* **22**, 1790-7 (2012).
6. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
7. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-43 (2013).
8. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98 (2012).
9. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).
10. He, B., Chen, C., Teng, L. & Tan, K. Global view of enhancer-promoter interactome in human cells. *Proc Natl Acad Sci U S A* **111**, E2191-9 (2014).
11. Corradin, O. *et al.* Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res* **24**, 1-13 (2014).
12. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
13. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-61 (2014).
14. Blankenberg, D. *et al.* Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* **Chapter 19**, Unit 19 10 1-21 (2010).
15. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
16. Amos, C.I. *et al.* The OncoArray Consortium: a Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol Biomarkers Prev* (in press).
17. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
18. Li, Y. *et al.* FastPop: a rapid principal component derived method to infer intercontinental ancestry using genetic data. *BMC Bioinformatics* **17**, 122 (2016).
19. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
20. Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* **33**, D428-32 (2005).
21. Schaefer, C.F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res* **37**, D674-9 (2009).
22. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).
23. Romero, P. *et al.* Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol* **6**, R2 (2005).
24. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
25. Kandasamy, K. *et al.* NetPath: a public resource of curated signal transduction pathways. *Genome Biol* **11**, R3 (2010).
26. Thomas, P.D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* **13**, 2129-41 (2003).

27. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
28. Kar, S.P. *et al.* Genome-Wide Meta-Analyses of Breast, Ovarian, and Prostate Cancer Association Studies Identify Multiple New Susceptibility Loci Shared by at Least Two Cancer Types. *Cancer Discov* **6**, 1052-67 (2016).
29. Braun, R. & Buetow, K. Pathways of distinction analysis: a new technique for multi-SNP analysis of GWAS data. *PLoS Genet* **7**, e1002101 (2011).
30. Jia, P., Zheng, S., Long, J., Zheng, W. & Zhao, Z. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics* **27**, 95-102 (2011).
31. Mogushi, K. & Tanaka, H. PathAct: a novel method for pathway analysis using gene expression profiles. *Bioinformation* **9**, 394-400 (2013).
32. Medina, I. *et al.* Gene set-based analysis of polymorphisms: finding pathways or biological processes associated to traits in genome-wide association studies. *Nucleic Acids Res* **37**, W340-4 (2009).