Genome-wide association analysis identifies three new breast cancer susceptibility loci

Breast cancer is the most common cancer among women. To date, 22 common breast cancer susceptibility loci have been identified accounting for ~8% of the heritability of the disease. We attempted to replicate 72 promising associations from two independent genome-wide association studies (GWAS) in ~70,000 cases and ~68,000 controls from 41 case-control studies and 9 breast cancer GWAS. We identified three new breast cancer risk loci at 12p11 (rs10771399; \( P = 2.7 \times 10^{-35} \)), 12q24 (rs1292011; \( P = 4.3 \times 10^{-19} \)) and 21q21 (rs2823093; \( P = 1.1 \times 10^{-12} \)). rs10771399 was associated with similar relative risks for both estrogen receptor (ER)-negative and ER-positive breast cancer, whereas the other two loci were associated only with ER-positive disease. Two of the loci lie in regions that contain strong plausible candidate genes: PTHLH (12p11) has a crucial role in mammary gland development and the establishment of bone metastasis in breast cancer, and NRIP1 (21q21) encodes an ER cofactor and has a role in the regulation of breast cancer cell growth.

Breast cancer is one of the most commonly occurring epithelial malignancies in women, with an estimated 1 million new cases and over 400,000 deaths annually worldwide. Familial aggregation and twin studies have shown the substantial contribution of inherited susceptibility to breast cancer. Over the last 4 years, we and others have conducted several GWAS and reported breast cancer susceptibility variants at 21 loci, with an additional locus (CASPI9) identified through a candidate gene approach. These variants are associated with modest risks for the disease (per-allele odds ratios (ORs) of <1.5) and explain ~8% of the excess familial risk of breast cancer, whereas other rarer high- and moderate-risk loci contribute less than 20%, suggesting that other loci remain to be identified.

To identify additional breast cancer susceptibility loci, we selected 72 SNPs for analysis that were genotyped and found to be significantly associated with breast cancer at \( P < 0.0001 \) in either of two breast cancer GWAS in the UK (UK2 and British Breast Cancer Study (BBCS)) and colon control studies through the Breast Cancer Association Consortium (BCAC). After quality control exclusions (see Online Methods), we analyzed data on 54,588 cases of invasive breast cancer, 2,401 cases of ductal carcinoma in situ (DCIS) and 58,998 controls. In addition, we used data from seven additional breast cancer GWAS from which summary results had been obtained on the basis of imputation to the HapMap 2 Utah residents of Northern and Western European ancestry (CEU) population. Results from the GWAS and BCAC replication were then combined to derive the overall evidence of association for each SNP based on 69,564 cases and 68,150 controls.

Three SNPs showed strong evidence for association in European women, consistent with the effect seen in the original GWAS (Fig. 1 and Table 1). In each case, the genotype-specific ORs were consistent with an allele dosage (log-additive) model (Supplementary Table 1). The rs2823093 SNP showed some evidence of heterogeneity in the per-allele ORs among studies in the replication stage (\( P = 0.002 \)), with particularly marked associations in two studies (Hannover-Minsk Breast Cancer Study (HMBCS) and Rotterdam Breast Cancer Study (RBCS); Fig. 1). The association in the replication stage remained highly significant, however, even after exclusion of these two studies (\( P = 7.1 \times 10^{-7} \)). The other two loci showed no evidence of heterogeneity among studies. Two additional SNPs at 17q21, rs2532348 and rs199523 (correlated at \( r^2 = 0.80 \) in the UK2 GWAS), gave more limited evidence of replication (\( P = 7.8 \times 10^{-5} \) and 0.0063, respectively) and reached \( P = 5.8 \times 10^{-7} \) and \( 2.6 \times 10^{-6} \), respectively, when combined with the GWAS data (Supplementary Table 2). These SNPs were only genotyped in the UK2 GWAS. They could not be imputed using HapMap and were only successfully genotyped in 12 studies in the BCAC replication. Moreover, for rs2532348, there was evidence of heterogeneity among studies in the per-allele ORs in BCAC (\( P = 0.001 \)). Additional data will be required to determine whether this SNP is associated with breast cancer risk. Three other SNPs (rs10940235 at 5q11, rs4403040 at 4q21 and rs6027564 at 2q13) showed evidence of replication at \( P < 0.01 \), but none reached genome-wide levels of statistical significance (Supplementary Table 2).

For women of Asian ancestry, rs10771399 (12p11) was also associated with breast cancer risk, with the estimated OR being similar to that in women of European ancestry (Supplementary Table 3). There was no significant evidence of association for either rs1292011 (12q24) or rs2823093 (21q21) in women of Asian ancestry. For rs2823093, the estimated OR was in the opposite direction from that in women of European ancestry, but the estimates did not differ significantly (Supplementary Table 3).

The rs10771399 SNP showed strong evidence of association with both ER-positive and ER-negative breast cancer, with the estimated per-allele ORs being similar (based on 24,775 ER-positive and 7,122 ER-negative cases; Supplementary Table 4a). In contrast, for rs1292011 and rs2823093, the association was confined to ER-positive breast cancer, with no evidence of association for ER-negative disease (Supplementary Table 4a). These latter results conform to the general pattern of a preponderance of common susceptibility loci containing weak and moderate-risk loci with modest risks for the disease (per-allele ORs of <1.3) and explain ~8% of the excess familial risk of breast cancer, whereas other rarer high- and moderate-risk loci contribute less than 20%, suggesting that other loci remain to be identified.

A full list of authors and affiliations appears at the end of the paper.

Received 16 August 2011; accepted 1 December 2011; published online 22 January 2012; doi:10.1038/ng.1049
for ER-positive disease identified through GWAS based on cases not selected for disease subtype\(^{19,20}\). In terms of per-allele OR, rs10771399 had one of the strongest effects identified to date for ER-negative breast cancer (OR = 0.85, 95% confidence interval (CI) = 0.80–0.90). For all three SNPs, the per-allele ORs for DCIS was similar to those for invasive disease (based on up to 2,148 DCIS cases; Supplementary Table 4b). For rs10771399, the estimated OR was higher for ten studies in which cases were selected for a positive family history and/or bilateral- ity, as would be expected under a polygenic model\(^{21}\) (\(P = 0.027\); Supplementary Table 5); however, exclusion of data from these studies made little difference to the estimated OR. There was no evidence for difference in the per-allele OR by age at diagnosis for any SNP (Supplementary Table 4c).

The rs10771399 SNP lies in an ~300-kb linkage disequilibrium (LD) block at 12p11 that contains one known gene, \(PTHLH\) (encoding parathyroid hormone–like hormone isofrm 1), also called \(PTHRp\) (parathyroid hormone–related protein) (Fig. 2a). \(PTHLH\) is expressed in a wide variety of tissues and in many malignancies, including 60% of breast tumors, and is required for normal mammary gland and bone development\(^{22–25}\). During lactation, PTHrP is released by the mammary gland to regulate the transfer of calcium from the skeleton to the milk\(^{26,27}\). Tumor-secreted PTHrP mimics the action of parathyroid hormone (PTH) by binding to its receptor \(PTHR1\)\(^{28}\), thereby promoting humoral hypercalcemia as well as metastasis of breast cancer cells to the bone\(^{29–31}\). It has been suggested that PTHrP enhances tumorigenesis through its pro-proliferative and anti-apoptotic activities by promoting survival in cells subjected to apoptosis\(^{32,33}\). However, conflicting data regarding the correlation of \(PTHLH\) expression levels and breast cancer survival have been reported\(^{34–36}\). Moreover, a recent study found that loss of PTHrP accelerates tumor incidence in DCIS and is associated with monocyte infiltration\(^{37}\).

**Table 1** Summary estimates for three loci associated with breast cancer risk

| SNP | Chromosome position* | Alleles* | MAF  | Stage          | Per-allele OR (95% CI)  | \(P\)  | Combined \(P\) |
|-----|----------------------|---------|------|----------------|-------------------------|-------|----------------|
| rs10771399 | 12p11 | A/G | 0.12 | UK2 | 0.79 (0.71–0.87) | 3.1 × 10\(^{-6}\) |
|       | 26,046,347          |         |     | Other GWAS | 0.83 (0.75–0.91) | 5.7 × 10\(^{-5}\) |
|       |                     | 0.11    |     | BCAC replication | 0.85 (0.83–0.88) | 3.3 × 10\(^{-27}\) |
|       | rs1292011            | A/G | 0.41 | UK2 | 0.92 (0.91–0.94) | 6.2 × 10\(^{-14}\) |
|       | 12q24                |         |     | BCAC replication | 0.92 (0.91–0.94) | 6.2 × 10\(^{-14}\) |
|       | 114,320,905          |         |     | Other GWAS | 0.91 (0.86–0.96) | 0.0008 |
|       | rs2823093            | G/A | 0.26 | UK2 | 0.96 (0.89–1.03) | 0.21  |
|       | 21q21                |         |     | BCAC replication | 0.94 (0.92–0.96) | 1.7 × 10\(^{-9}\) |
|       | 15,442,703           |         |     | Other GWAS | 0.91 (0.85–0.97) | 0.0032 |
|       |                     | 0.26    |     | BCAC replication | 0.94 (0.92–0.96) | 1.7 × 10\(^{-9}\) |

*Build 36. †Minor allele listed second. ‡Per copy of the minor allele.
The rs1292011 SNP at 12q24 lies in an ~100-kb LD block that contains no known genes (Fig. 2b). SNPs in this region have been found to be associated with squamous esophageal carcinoma, renal cell carcinoma, liver adenoma, heart disease and type 1 diabetes as well as blood pressure and prostate-specific antigen (PSA) levels. Two plausible cancer candidate genes, MAPKAPK5 (encoding mitogen-activated protein kinase–activated protein kinase 5; also called MK5 or PRAK) and TBX3 (encoding T-box 3), lie within 2 Mb of rs1292011. The MAPKAPK5 protein is a member of the serine/threonine kinase family, and MAPKAPK5 is directly activated by Myc. TBX3 has a role in mammary gland development and haploinsufficiency for TBX3 is associated with unlar-mammary disorder. TBX3 was found to be amplified and overexpressed in several cancers, including breast cancer, and TBX3 was at high levels in plasma from individuals with breast and ovarian cancer. Recently, it has been shown that estrogen regulates the expansion of breast cancer stem cells through the fibroblast growth factor (FGF)–FGF receptor (FGFR)-TBX3 pathway and that TBX3 is a direct downstream target of the Wnt–β-catenin pathway. The expression of TBX3 was found to be significantly higher (P < 0.0001) in ER-positive breast cancer tumors than in ER-negative tumors in two independent data sets containing 781 tumors (using expression data from the Human Genome HG 133A Affymetrix array) and 244 tumors (using expression data from the Human Agilent 44k microarray). These data suggest that the association of rs1292011 with ER-positive breast cancer could be mediated through its effect on TBX3.

The rs2823093 SNP lies in an ~130-kb LD block containing no known genes. The nearest gene, ~900 kb downstream, is NRIP1 (encoding nuclear receptor–interacting protein 1) (Fig. 2c), also called RIP140 (receptor-interacting protein 140). RIP140 acts as a strong transcriptional repressor for nuclear receptors. It interacts with estrogen receptor α (ERα), represses ER signaling and inhibits its mitogenic effects. This repression is mediated through interaction with FHL1, a protein involved in suppressing cancer cell growth and migration. Several lines of evidence suggest that RIP140 has an important role in the regulation of breast cancer cell growth. Knockdown of RIP140 was found to induce growth promotion in an ER-positive breast cancer cell line. The RIP140 protein was also highly induced following the treatment of human breast cancer cells with retinoids, known for their breast cancer growth suppression and anti-estrogenic effects.

A Spanish case-control study, which genotyped SNPs in 91 breast cancer candidate genes in ~700 cases and ~700 controls, identified a relatively rare SNP at this locus (rs926184; minor allele frequency (MAF) of ~2%) located 175 kb upstream of rs2823093, which showed a modest association with breast cancer. These two SNPs, however, are not correlated (r² = 0 in the HapMap CEU population). The expression of NRIP1 has been shown to be significantly higher in ER-positive tumors relative to ER-negative ones (P < 0.0001), suggesting that the association of rs2823093 with ER-positive breast cancer could be mediated through its effect on NRIP1 expression.

The three susceptibility variants newly identified in this study are relatively common (MAFs of 0.11–0.41) and together explain ~0.7% of the familial risk of breast cancer and bring the total contribution of common low-penetrance breast cancer susceptibility loci to ~9%. The relative risks associated with these variants are modest, with the per-allele ORs for the risk alleles ranging from 1.07–1.22-fold, but the causal variants underlying some of these loci might confer more substantial risks. The present work highlights the importance of combining GWAS and large-scale replication studies with tumor subtyping in the identification and characterization of breast cancer susceptibility loci.

**Figure 2** Association plots for the three new breast cancer susceptibility loci drawn using SNP annotation and proxy search (SNAP) software. Genotyped and imputed SNPs are plotted based on their chromosomal position in build 36 on the x axis and their overall P values (−log₁₀ values) from the UK2 and BBCS GWAS on the y axis. For each region, the most strongly associated SNP is represented by a diamond. The intensity of the red shading reflects the strength of correlation (r²) between the best SNP and the other SNPs in the region. Genes present in the region (if any) are indicated in green. Data are shown for (a) 12p11, (b) 12q24 and (c) 21q21.
The genes in these regions (if proven to be the causal genes) under-score that diverse mechanisms are likely to be relevant to breast cancer pathogenesis. Re-sequencing of these loci, combined with fine-scale mapping and functional analyses, will provide greater insights into the genetic architecture of breast cancer and the pathogenesis of the disease.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the following individuals for their contribution to this project (study names given in parentheses): S. Cornelissen, R. van Hien, L. Braaf, L. van Veer, B. Bueno-de-Mesquita and S. Canisius (ABCs); N. McNieiner, G. Colleran, A. Rowan and N. Miller (BIGGS); A. Langbeisz (BSUCH); J. Arias, P. Zamora, P. Menendez, T. Moreno and G. Pita (CNO-BCS); M. Adank, M. Ausems and S. Verhoeof (DFBCS); U. Hamann, Y.-D. Ko, C. Baisch, H.-P. Fischer, B. Pesch, S. Rabenstein and V. Harth (GENICA); K. Aaltosen, P. Heikkila, T. Heikkinen, D. Greco, R.N.H. Jantti and I. Erkkila (HEBCS); H. Kemiläinen, E. Myöhänien and A. Parkkiniemi (KBCP); T. Slinger, E. Mutschelknauss, S. Behrens, R. Birr, M. Celik, U. Eilber, B. Kaspereit, N. Knese and K. Smit (MARIE); P. Radice, B. Peisell, M. Barle and M.A. Pierotti (MBCSG); T. Selander, M. Gill, L. Collins and N. Weerbaarsoy (OFBCR); M. Gripp, K. Mononen and M. Otsuoka (OBBCS); E. Krol-Warmerdam and J. Blom (ORIGO); and P. Boonyawongviroj and P. Sirivesanarungsan (ACP). For full acknowledgments, including funding sources, see the Supplementary Note.

AUTHOR CONTRIBUTIONS

M.G. and D.F.E. wrote the manuscript. K.M., M.G. and D.F.E. performed the statistical analysis. O.F., N.I., O.L., I.S.L.S., M.L. and J.P. led the BCAS GWAS. D.F.E., P.D.P.P., A.M.D., C.T. and N.R. led the UK2 GWAS. Q. Waisfisz and H.M.-H. led the DFBCS GWAS, with support from A.G.U. and F.R., P. Hall, K.C., A.I. and J. Liu led the SASBAC GWAS. H.M., N.R. and C. Seynaeve provided data management support for BCAC. C.J.-C., R.H., S.N. and D.F.-J. led the MARIE GW AS. J.L.H., M. Southey, H.T., E.M., D.F.E., W.J.T., S.M.G. and N.J.G. coordinated the POSH study. D.M.E., W.J.T., S.M.G. and N.J.G. coordinated the PSBCH study. D.M.E., W.J.T., S.M.G. and N.J.G. coordinated the PSH study. HEBON organized patient recruitment and data and sample collection for the DFBBCS GWAS. FBCS organized patient recruitment and data and sample collection for the UK2 GWAS. The GENICA Network organized patient recruitment and data and sample collection for the GENICA study. kConFab investigators and the Australian Ovarian Cancer Study organized recruitment and data and sample collection for KConFab and AOCs, respectively.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Kamangar, F., Dores, G.M. & Anderson, W.F. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. J. Clin. Oncol. 24, 2137–2150 (2006).

2. Lichtenstein, P. et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. N. Engl. J. Med. 343, 78–85 (2000).

3. Petro, J. & Mack, T.M. High constant incidence in twins and other relatives of women with breast cancer. Nat. Genet. 26, 411–414 (2000).

4. Ahmed, S. et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat. Genet. 41, 585–590 (2009).

5. Antoniou, A.C. et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor–negative breast cancer in the general population. Nat. Genet. 42, 885–892 (2010).

6. Easton, D.F. et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447, 1087–1093 (2007).

7. Stacey, S.N. et al. Novel breast cancer susceptibility loci at 9q31.2: results of a genome-wide association study. J. Natl. Cancer Inst. 103, 425–435 (2011).

8. Haiman, C.A. et al. A common variant at the TERT–CLPTM1L locus is associated with estrogen receptor–negative breast cancer. Nat. Genet. 43, 1210–1214 (2011).

9. Hunter, D.J. et al. A genome-wide association study identifies alleles in FGR2 associated with risk of sporadic postmenopausal breast cancer. Nat. Genet. 39, 870–874 (2007).

10. Stacey, S.N. et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen-receptor-positive breast cancer. Nat. Genet. 39, 865–869 (2007).

11. Stacey, S.N. et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor–positive breast cancer. Nat. Genet. 40, 703–706 (2008).

12. Thomas, G. et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat. Genet. 43, 679–684 (2009).

13. Turnbull, C. et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat. Genet. 42, 504–507 (2010).

14. Zheng, W. et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q22. J. Natl. Cancer Inst. 103, 324–328 (2011).

15. Cox, A. et al. A common coding variant in CASP8 is associated with breast cancer risk. Nat. Genet. 39, 352–358 (2007).

16. Varghese, J.S. & Easton, D.F. Genome-wide association studies in common cancers—what have we learnt? Curr. Opin. Genet. Dev. 20, 201–209 (2010).

17. Fletcher, O. et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. J. Natl. Cancer Inst. 103, 425–435 (2011).

18. Turnbull, C. et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat. Genet. 42, 504–507 (2010).

19. Brooks, A. et al. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. Hum. Mol. Genet. 20, 3289–3303 (2011).

20. Garcia-Closas, M. et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathologcal characteristics. PLoS Genet. 4, e1000054 (2008).

21. Antoniou, A.C. & Easton, D.F. Risk prediction models for familial breast cancer. Future Oncol. 2, 257–274 (2006).

22. Durban, M.E., Wysolmerski, J.J. & Brodux, A.E. Parathyroid hormone-related protein: from hypercalcemia of malignancy to developmental regulatory molecule. Annu. Rev. Physiol. 60, 431–460 (1998).

NATURE GENETICS VOLUME 44 | NUMBER 3 | MARCH 2012
26. Kovacs, C.S. & Kronenberg, H.M. Maternal-fetal calcium and bone metabolism during pregnancy, puerperium, and lactation. Endocr. Rev. 18, 832–872 (1997).

27. Wysolmerski, J.J. Interactions between breast, bone, and brain regulate mineral and skeletal metabolism during lactation. Ann. NY Acad. Sci. 1192, 161–169 (2010).

28. Jüppner, H. et al. A G protein–linked receptor for parathyroid hormone and parathyroid hormone–related peptide. Science 254, 1024–1026 (1991).

29. DeMauro, S. & Wysolmerski, J. Hypercalcaemia in breast cancer: an echo of bone mobilization during lactation? J. Mammary Gland Biol. Neoplasia 10, 157–167 (2005).

30. Durcan M.E. & Wysolmerski, J.J. Parathyroid hormone–related protein: a developmental regulatory molecule necessary for mammary gland development. J. Mammary Gland Biol. Neoplasia 4, 21–34 (1999).

31. Wysolmerski, J.J. & Broadus, A.E. Hypercalcaemia of malignancy: the central role of parathyroid hormone–related protein. Annu. Rev. Med. 45, 189–200 (1994).

32. Henderson, J.E. et al. Nuclear localization of parathyroid hormone–related peptide enhances survival of chondrocytes under conditions that promote apoptotic cell death. Mol. Cell. Biol. 15, 4064–4075 (1995).

33. Okoumassou, L.E., Russo, C., Denizot, F., Avril-Bates, D. & Henderson, J.E. Parathyroid hormone–related protein (PTHrP) inhibits mitochondrial-dependent apoptosis through CK2. J. Cell. Physiol. 212, 591–599 (2007).

34. Hanson, M.A. et al. Parathyroid hormone–related protein localization in breast cancers predict improved prognosis. Cancer Res. 66, 2250–2256 (2006).

35. Linforth, R. et al. Coexpression of parathyroid hormone related protein and its receptor in early breast cancer predicts poor patient survival. Clin. Cancer Res. 8, 3172–3177 (2002).

36. Linforth, R. et al. Significance of the parathyroid hormone–related protein expression in breast carcinoma. Breast Cancer 7, 215–220 (2000).

37. Fleming, N.I. et al. Parathyroid hormone–related protein protects against mammary tumor emergence and is associated with monozygotic inheritance in ductal carcinoma in situ. Cancer Res. 69, 7473–7479 (2009).

38. Bluteau, O. et al. Bi-allelic inactivation of TCF1 in hepatic adenomas. Nat. Genet. 32, 312–315 (2002).

39. Cho, Y.S. et al. A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. Nat. Genet. 41, 527–534 (2009).

40. Cui, R. et al. Functional variants in ADH1B and ALDH2 coupled with alcohol and smoking synergistically enhance esophageal cancer risk. Gastroenterology 137, 1768–1775 (2009).

41. Erdmann, J. et al. New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat. Genet. 41, 280–282 (2009).

42. Gudbjartsson, D.F. et al. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. Nat. Genet. 41, 342–347 (2009).

43. Kato, N. et al. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. Nat. Genet. 43, 531–538 (2011).

44. Purcell, S. et al. Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. Nat. Genet. 43, 60–65 (2011).

45. Soranzo, N. et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat. Genet. 41, 1182–1190 (2009).

46. Todd, T.A. et al. Robust associations of four new chromosomal regions from genome-wide analyses of type 1 diabetes. Nat. Genet. 39, 857–864 (2007).

47. Wu, C. et al. Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. Nat. Genet. 43, 67–72 (2011).

48. Kress, T.R. et al. The MK5/PRAK kinase and Myc form a negative feedback loop inactivation during lactation? Mammary gland, limb and yolk sac defects in mice lacking Tbx3, the gene mutated in human uIAR syndromes. Development 130, 2263–2273 (2003).

49. Bamshad, M. et al. Mutations in human Tbx3 alter limb, apocrine and genital development in uIAR-mammary syndrome. Nat. Genet. 16, 311–315 (1997).

50. Fan, W., Huang, X., Chen, C., Gray, J. & Huang, T. Tbx3 and its isoform Tbx3+2a are functionally distinctive in inhibition of senescence and are overexpressed in a subset of breast cancer cell lines. Cancer Res. 64, 5132–5139 (2004).

51. Lomnitzka, M., Dubrowska, A., Helman, U., Volodko, N. & Souchelnytskyi, S. Increased expression of c5HMT, Tbx3 and utrophin in plasma of ovarian and breast cancer patients. Int. J. Cancer 118, 421–427 (2006).

52. Lyng, H. et al. Gene expressions and copy numbers associated with metastatic phenotypes of uterine cervical cancer. BMC Genomics 7, 268 (2006).

53. Rowley, M., Grotthey, E. & Couch, F.J. The role of Tbx2 and Tbx3 in mammary development and tumorigenesis. J. Mammary Gland Biol. Neoplasia 9, 109–118 (2004).

54. Fillmore, C.M. et al. Estrogen expands breast cancer stem-like cells through paracrine FGF/Tbx3 signaling. Proc. Natl. Acad. Sci. USA 107, 21737–21742 (2010).

55. Renard, C.A. et al. Tbx3b is a downstream target of the Wnt/b-catenin pathway and a critical mediator of b-catenin survival functions in liver cancer. Cancer Res. 67, 901–907 (2007).

56. Reyal, F. et al. A comprehensive analysis of prognostic signatures reveals the high predictive capacity of the proliferation, immune response and RNA splicing modules in breast cancer. Breast Cancer Res. 10, R93 (2008).

57. van de Vijver, M.J. et al. A gene expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. 347, 1999–2002 (2002).

58. L’Hortet, F., Daivou, S., Heery, D.M., Cavaillès, V. & Parker, M.G. RIP-140 interacts with multiple nuclear receptors by means of two distinct sites. Mol. Cell. Biol. 16, 6029–6036 (1996).

59. Tjirna, H. et al. Regulation of subnuclear localization is associated with a mechanism for nuclear receptor corepression by RIP140. Mol. Cell. Biol. 23, 4187–4198 (2003).

60. White, K.A., Yore, M.M., Deng, D. & Spinella, M.J. Limiting effects of RIP140 in estrogen signaling: potential mediation of anti-estrogenic effects of retinoid acid. J. Biol. Chem. 280, 7829–7835 (2005).

61. Lin, J. et al. Four and a half LIM domains 1 (FHL1) and receptor interacting protein of 140xOda (RIP140) interact and cooperate in estrogen signaling. Int. J. Biochem. Cell Biol. 41, 1613–1618 (2009).

62. Kerley, J.S., Olsen, S.L., Freemantle, S.J. & Spinella, M.J. Transcriptional activation of the nuclear receptor corepressor RIP140 by retinoic acid: a potential negative-feedback regulatory mechanism. Biochem. Biophys. Res. Commun. 285, 969–975 (2001).

63. White, K.A. et al. Negative feedback at the level of nuclear receptor coregulation. Self-limitation of retinoid signaling by RIP140. J. Biol. Chem. 278, 43889–43892 (2003).

64. Demerac, E. et al. Antiestrogenic effects of all-trans-retinoic acid and 1,25-dihydroxyvitamin D3 in breast cancer cells occur at the estrogen response element level but through different molecular mechanisms. Cancer Res. 54, 14584–14586 (1994).

65. Fontana, L.A., Nervi, C., Shao, Z.M. & Jetten, A.M. Retinoid antagonism of estrogen-response transforming growth factor α and β2 gene expression in breast carcinoma cells. Cancer Res. 52, 3938–3945 (1992).

66. Vega, A. et al. Evaluating new candidate SNPs as low penetrance risk factors in sporadic breast cancer: a two-stage Spanish case-control study. Gynecol. Oncol. 67(1), 210–214 (2000).

67. Johnson, A.D. et al. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24, 2938–2939 (2008)
Roger L Milne44, M Rosario Alonso45, Anna González-Neira45, Javier Benítez46, Hoda Anton-Culver47, Argyrios Ziogas47, Leslie Bernstein48, Christine Clarke Dur49, Hermann Brenner50, Heiko Müller50, Volker Arndt50, Christa Stegmaier51, Familial Breast Cancer Study (FBCS)4,10, Christina Justenhoven52,53, Hiltrud Brauch2,53, Thomas Brüning54, The Gene Environment Interaction of Breast Cancer in Germany (GENICA) Network50,52–56, Shan Wang-Gohrke57, Ursula Eibler58, Thilo Dörk58, Peter Schürmann58, Michael Bremer58, Peter Hillemans58, Natalia V Bogdanova59, Natalia N Antonenko60, Yuri I Rogo60, Johann H Karstens59, Marina Bermeheva61, Darya Prokofieva61, Elza Khusnutdinova61, Annika Lindblom62, Sara Margolin63, Arto Mannnermaa64–66, Vesa Kataja65,67,68, Veli-Matti Kosma64–66, Jaana M Hartikainen64–66, Dietmar Lambrecht69, Betul T Yesilyurt70, Giuseppe Floris71, Karin Leunen70, Stanouš Manoukian71, Bernardo Bonanni72, Stefano Fortuzz73, Paolo Peterlongo73,74, Fergus J Couch75,76, Xianshu Wang75, Kristen Stevens76, Adam Lee77, Graham G Giles75,78, Laura Baglietto75,78, Gianluca Severi75,78, Katrinna McLean79, Grethe Grenaker Alnæs80, Vesselia Kristensen80,81, Anne-Lise Børresen-Dale80,81, Esther M John9,82, Alexander Miron83, Robert Winqvist84, Kari Pylkäs84, Arja Jukkola-Vuorinen85, Saila Kaupilla86, Irene L Andrulis87,88, Gerd Glendon89, Anna Marie Mulligan90,91, Peter Devilee92,93, Christie J van Asperen94, Rob A E M Tollenaar95, Caroline Seynaeve96, Jonine D Figueroa27, Montserrat Garcia-Closas4,97, Louise Brinton27, Jolanta Lisowska98, Maartje Hooning99, Antoinette Hollestelle100, Rogier A Oldenburg101, Ans M W van den Oudewal101, Angela Cox102, Malcolm W Reed103, Mitul Shah1, Ania Jakubowska104, Jan Lubinski104, Katarzyna Jaworska104, Katarzyna Durda105, Michael Jones97, Minouk Schoemaker97, Alan Ashworth10, Anthony Swerdlov97, Jonathan Beesley105, Xiaoxiao Chen105, K-ConFab Investigator10, Australian Ovarian Cancer Study Group10, Kenneth R Muir106, Arttaya Lophatananon106, Suthee Rattanamongkong107, Arkom Chaiverawatana108, Daech Kang109, Keun-Young Yoo109, Dong-Young Noh109, Chen-Yang Shen110, Jyh-Cherng Yu111, Pei-Ei Wu110, Chia-Ni Hsiung110, Annie Perkins111, Ruth Swann112, Louiza Velentzis112, Diana M Eccles113, Will J Tapper113, Susan M Gerty113, Nikki J Graham113, Bruce A J Ponder114,115, Georgia Chenevix-Trench115, Paul D P Pharoah1,13, Mark Lathrop116,117, Alison M Dunning1, Nazneen Rahman4, Julian Peto7 & Douglas F Easton1,13

1Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. 2Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London, UK. 3Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. 4Section of Cancer Genetics, Institute of Cancer Research, Sutton, UK. 5Department of Experimental Therapy, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. 6Department of Epidemiology, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. 7Non-Communicable Disease Epidemiology Department, London School of Hygiene and Tropical Medicine, London, UK. 8Department of Clinical Genetics, Section Oncogenetics, VU University Medical Center, Amsterdam, The Netherlands. 9Department of Internal Medicine and Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. 10Full list of contributing authors is provided in the Supplementary Note. 11Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. 12Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. 13Human Genetics Division, Genome Institute of Singapore, Singapore. 14Department of Clinical Genetics, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. 15Department of Oncology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. 16Clinic of Gynaecology and Obstetrics, Division for Gynaecological Tumor Genetics, Technische Universität München, Munich, Germany. 17Department of Obstetrics and Gynecology, Division of Molecular Gynaeco-Oncology, University of Cologne, Cologne, Germany. 18Max Planck Institute of Psychiatry, Munich, Germany. 19Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. 20Division of Cancer Epidemiology, Deutsches Krebsforschungszentrum, Heidelberg, Germany. 21Primär Medizinische Versorgung (PMV) Research Group at the Department of Child and Adolescent Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany. 22Department of Cancer Epidemiology/Clinical Cancer Registry, University Clinic Hamburg-Eppendorf, Hamburg, Germany. 23Institute for Medical Biometrics and Epidemiology, University Clinic Hamburg-Eppendorf, Hamburg, Germany. 24Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia. 25Centre for Molecular, Environmental, Genetic, and Analytic Epidemiology, Melbourne School of Population Health, The University of Melbourne, Melbourne, Victoria, Australia. 26Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. 27Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA. 28Department of Clinical Genetics, Family Cancer Clinic, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. 29Department of Molecular Pathology, Family Cancer Clinic, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands. 30Department of Gynecology and Obstetrics, University Breast Center, University Hospital Erlangen, Erlangen, Germany. 31Institute of Human Genetics, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany. 32Division of Cancer Studies, National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre, Guy’s & St. Thomas’ National Health Service (NHS) Foundation Trust, London, UK. 33Division of Cancer Studies, King’s College London, London, UK. 34Wellcome Trust Centre for Human Genetics and Oxford Biomedical Research Centre, University of Oxford, Oxford, UK. 35Clinical Science Institute, University Hospital Galway, Galway, Ireland. 36Department of Obstetrics and Gynecology, University of Heidelberg, Heidelberg, Germany. 37National Center for Tumor Diseases, University of Heidelberg, Heidelberg, Germany. 38Molecular Epidemiology Unit, German Cancer Research Center (DFKZ), Heidelberg, Germany. 39Institut National de la Santé et la Recherche Médicale (Inserm), Centre for Research in Epidemiology and Population Health (CESP), U1018, Environmental Epidemiology of Cancer, Villejuif, France. 40Université Paris-Sud, Unité Mixte de Recherche en Santé (UMRS) 1018, Villejuif, France. 41Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital, Copenhagen, Denmark. 42Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Copenhagen, Denmark. 43Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Copenhagen, Denmark. 44Genetic & Molecular Epidemiology Group, Human Cancer Genetics Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. 45Human Genotyping Unit, Human Cancer Genetics Programme, CNIO, Madrid, Spain. 46Department of Epidemiology, University of California–Irvine, Irvine, California, USA. 47City of Hope Cancer Centre, Duarte, California, USA. 48Cancer Prevention Institute of California, Fremont, California, USA. 49Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.
ONLINE METHODS

GWAS analysis. Primary genotype data were obtained for nine breast cancer GWAS in populations of European ancestry (Supplementary Table 6). Standard quality control analysis was performed on all scans. We excluded all individuals with low call rate (<95%) or extremely high or low heterozygosity ($P < 1 \times 10^{-3}$) and all individuals determined to be of non-European ancestry (>15% non-European component, by multidimensional scaling using the CEU, Yoruba in Ibadan (YRI), Han Chinese in Beijing (CHB) and Japanese in Tokyo (JPT) HapMap 2 populations as a reference). We excluded SNPs with a call rate of <95%, a call rate of <99% and a MAF of <5%, all SNPs with a MAF of <1% and SNPs whose genotype frequencies departed from Hardy-Weinberg equilibrium at $P < 1 \times 10^{-6}$ in controls or $P < 1 \times 10^{-12}$ in cases. For highly significant SNPs, the genotype intensity cluster plots were examined manually to judge reliability, either centrally or by contacting the original investigators.

Data were imputed for all scans for ~2.6 million SNPs using the HapMap 2 CEU population as a reference, using the program MACH v1.0. Estimated per-allele ORs and standard errors were generated from the imputed genotypes using Probabel\(^6^\). For two studies (UK2 and HEBCS), estimates were adjusted by the first three principal components, as this was found to materially reduce the inflation. We then adjusted for residual inflation by multiplying the variance by a genomic control adjustment factor, based on the ratio of the median $z$-squared test statistic to its expected value. The BBCS and UK2 studies used the same control data (Wellcome Trust Case Control Consortium 2; WTCCC2) but different genotyping platforms. These studies were imputed separately. For the combined analysis, the control set was divided randomly between the two studies, in proportion to the size of the case series, to provide disjoint strata. For a limited subset of SNPs that could not be imputed (including rs2532348 and rs195523 on 17q21), genotype data from the original scan(s) were used in the analysis.

Replication stage. SNPs for replication were genotyped in 46 studies, of which 4 were case-only studies that did not contribute to the current analysis (Supplementary Table 7). Data from BBCS were excluded, as the same cases were included in the GWAS. Seven studies (HABCS, HMBCS, HUBCS, KARBAC, RBCS, SEARCH and SEBCS) were analyzed by Fluidigm for 72 SNPs (Supplementary Table 2). We selected 63 SNPs from the UK2 study: one was replaced by a better surrogate and one failed, such that data were only available for 61 SNPs. Ten SNPs were selected from the BBCS study, and one SNP was selected from both scans (the original SNP, rs1973930, also referred to as rs5603999, did not work by Fluidigm and in some iPLEX analyses and was replaced by the surrogate rs10771399 ($r^2 = 0.95$), which was genotyped in all studies). Samples from 27 studies were genotyped by iPLEX for 29 SNPs that showed the strongest associations. Seven additional studies (ABBCS, CGPS, MCCS, NC-BCFR, OFBCR, PBCS and UKBG5) were genotyped by TaqMan for up to four SNPs that showed association after the Fluidigm and iPLEX genotyping, including all three SNPs discussed in detail here. We restricted the analysis to individuals of European or east Asian ancestry, as the sample size for other ancestral groups was too small to give meaningful results.

All studies were complied with BCAC genotyping quality control standards by including at least 2% of samples in duplicate and a common set of 93 Utah residents of Northern and Western European ancestry (CEPH) DNA sequences from the HapMap Consortium. Genotype data were excluded for any sample that consistently failed genotyping for >20% of the SNPs genotyped, all samples on any one plate that had an SNP call rate of <90%, all genotype data for any SNP where the overall call rate was <95% and all genotype data for any SNP where duplicate concordance was <98% (based on 2% of samples genotyped in duplicate). In addition, for any SNP for which the $P$ value for departure from Hardy-Weinberg equilibrium for controls was <0.005, clustering of the intensity plots was reviewed manually and the data excluded if clustering was judged to be poor. After quality control exclusions, we analyzed data on 54,588 cases of invasive breast cancer, 2,401 cases of DCIS and 58,098 controls.

Per-allele and genotype-specific ORs for the replication stage were estimated using logistic regression, adjusted for the study. Women of European and Asian ancestry were analyzed separately. NC-BCFR contributed cases and controls to both European and Asian analyses; for the remaining studies, the subjects were either predominantly European or predominantly Asian, and subjects from other minority ancestral groups were excluded. Statistical significance levels from the GWAS and BCAC replication phases were obtained by combining the logOR estimates and standard errors as in a fixed-effect meta-analysis. Heterogeneity in the OR association with each SNP by ER status was evaluated using a case-only analysis by logistic regression. Heterogeneity by age was evaluated by fitting a linear age x genotype interaction term. All participants provided written informed consent. Ethical approval for each study is described in Supplementary Table 7.

69. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 11, 134 (2010).