Genome-wide association study identifies a common variant in \textit{RAD51B} associated with male breast cancer risk

Nick Orr$^{1,2}$, Alina Lemnrau$^{1,2}$, Rosie Cooke$^{1,3}$, Olivia Fletcher$^{1,2}$, Katarzyna Tomczyk$^{1,2}$, Michael Jones$^{1,2}$, Nichola Johnson$^{1,2}$, Christopher J Lord$^{1,2}$, Costas Mitsopoulos$^{1,2}$, Marketa Zvelebil$^{1,2}$, Simon S McDade$^{4}$, Gemma Buck$^{3}$, Christine Blancher$^{5}$, KConFab Consortium$^{6,7}$, Alison H Trainer$^{8}$, Paul A James$^{8}$, Stig E Bojesen$^{9}$, Kathleen Cuningham Consortium for Research into Familial Breast Cancer (kConFab), Research Department, Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia. 2Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, UK. 3Centre for Cancer Research and Cell Biology, School of Medicine, Dentistry and Biomedical Science, Queen's University of Belfast, Belfast, UK. 4High-Throughput Genomics Group, Wellcome Trust Centre for Human Genetics, Oxford, UK. 5Familial Cancer Clinic, Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia. 6ISPO, Florence, Italy. 7Section of Epidemiology, with validation in independent sample sets totaling 438 cases and 474 controls. A SNP in \textit{RAD51B} at 14q24.1 was significantly associated with male breast cancer risk ($P = 3.02 \times 10^{-13}$; odds ratio (OR) = 1.57). We also refine association at 16q12.1 to a SNP within \textit{TOX3} ($P = 3.87 \times 10^{-15}$; OR = 1.50).

Male breast cancer accounts for 1% of all breast cancer diagnoses. Family history is a significant risk factor for male breast cancer: the relative risk of breast cancer for a female with an affected brother is approximately 30% higher than for a female with an affected sister. Approximately 10% of male breast cancer cases carry \textit{BRCA2} mutations, whereas \textit{BRCA1} mutation carriers are reported less frequently. We performed a genome-wide association study (GWAS) for male breast cancer, predicated on the assumption that common variation contributes appreciably to the heritable risk of male breast cancer and because investigation of risk alleles for breast cancer in males may provide insight into genetic susceptibility for the disease in females.

Using Illumina OmniExpress arrays, we genotyped 920 male breast cancer cases ascertained from the UK ($n = 805$) and United States ($n = 115$) (Supplementary Table 1 and Supplementary Methods). For controls, we used publicly available data on 2,912 individuals from the 1958 British Birth Cohort, genotyped on Illumina 1.2M Duo custom arrays. After applying prespecified quality control measures (Supplementary Fig. 1, Supplementary Table 2a,b and Supplementary Methods), we estimated ORs and 95% confidence intervals (CIs) for 447,760 autosomal SNPs with minor allele frequencies (MAFs) of 25% in 823 cases and 2,795 controls. Quantile-quantile plots of $P$ values showed minimal inflation of test statistics, indicating that there was not substantial cryptic population substructure or differential genotyping between cases and controls (genomic control inflation factor $\lambda = 1.05$; Supplementary Fig. 2).

A total of 17 SNPs, mapping to 6 independent genomic regions, showed evidence of association with male breast cancer at $P \leq 5.0 \times 10^{-7}$ (Supplementary Fig. 3). We attempted to validate the most significantly associated SNP mapping to each of the 6 regions in 438 cases and 474 controls recruited from 12 case-control series (Supplementary Table 1 and Supplementary Methods). In a combined analysis, the associations of two SNPs, rs1314913 ($P = 3.02 \times 10^{-13}$; OR = 1.57) and rs3803662 ($P = 3.87 \times 10^{-15}$; OR = 1.50), attained genome-wide significance (Table 1 and Supplementary Tables 3 and 4).

The rs1314913 SNP is located in intron 7 of the \textit{RAD51B} gene (encoding RAD51 homolog B) on chromosome 14q24.1 at 67,769,347 bp (NCBI Build 36). It maps to the distal end of a linkage

\begin{itemize}
  \item 1The Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK.
  \item 2Division of Breast Cancer Research, The Institute of Cancer Research, Sutton, UK.
  \item 3Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, UK.
  \item 4Centre for Cancer Research and Cell Biology, School of Medicine, Dentistry and Biomedical Science, Queen’s University of Belfast, Belfast, UK.
  \item 5High-Throughput Genomics Group, Wellcome Trust Centre for Human Genetics, Oxford, UK.
  \item 6Kathleen Cunningham Consortium for Research into Familial Breast Cancer (kConFab), Research Department, Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia.
  \item 7A list of members and affiliations is provided in the Supplementary Note.
  \item 8Familial Cancer Clinic, Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia.
  \item 9Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark.
  \item 10Department of Breast Surgery, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark.
  \item 11Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland.
  \item 12Department of Oncology, Helsinki University Central Hospital, Helsinki, Finland.
  \item 13The Susanne Levy Gertner Oncogenetics Unit, Sheba Medical Centre, Tel Aviv, Israel.
  \item 14The Sackler School of Medicine, Tel-Aviv University, Tel Aviv, Israel.
  \item 15Molecular and Nutritional Epidemiology Unit, Cancer Research and Prevention Institute (ISPO), Florence, Italy.
  \item 16Department of Molecular Medicine, Sapienza University of Rome, Rome, Italy.
  \item 17Department of Medical Oncology, Erasmus University Medical Center, Rotterdam, The Netherlands.
  \item 18Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands.
  \item 19Department of Molecular Diagnostics, Institute of Oncology Ljubljana, Ljubljana, Slovenia.
  \item 20Institute of Oncology Ljubljana, Cancer Genetics Clinic, Epidemiology and Cancer Registry, Ljubljana, Slovenia.
  \item 21Galician Foundation of Genome Medicine, Complejo Hospitalario Universitario Santiago (CHUS), Servicio Galego de Saúde (SERGAS), Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago de Compostela, Spain.
  \item 22Genetic Oncology Unit, Complejo Hospitalario Universitario de Vigo (CHUVI), SERGAS, Vigo, Spain.
  \item 23Department of Oncology, Lund University, Lund, Sweden.
  \item 24Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK.
  \item 25Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK.
  \item 26Section of Epidemiology and Biostatistics, Division of Cancer Studies and Pathology, Leeds Institute of Molecular Medicine, University of Leeds, Leeds, UK.
  \item 27Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, California, USA.
\end{itemize}

Received 3 April; accepted 29 August; published online 23 September 2012; doi:10.1038/ng.2417
Table 1  Summary data for rs1314913 at 14q24.1 and rs3803662 at 16q12.1 associated with risk of male breast cancer

| Locus    | Study stage | Control MAF | Control genotype counts | Case MAF | Case genotype counts | P value  | OR (95% CI) |
|----------|-------------|-------------|--------------------------|----------|----------------------|----------|-------------|
| rs1314913| GWAS        | 0.14        | GG, GA, AA               | 0.21     | GG, GA, AA           | 4.09 x 10^-10 | 1.55, 1.35–1.78 |
| 14q24.1  | Replication | 0.15        | 2.047, 675, 65           | 0.22     | 258, 155, 16         | 1.71 x 10^-4  | 1.61, 1.25–2.07 |
| Combined |             | 0.15        | 2.380, 782, 77           | 0.21     | 778, 416, 58         | 3.02 x 10^-13 | 1.57, 1.39–1.77 |
| rs3803662| GWAS        | 0.26        | GG, GA, AA               | 0.34     | GG, GA, AA           | 2.51 x 10^-10 | 1.46, 1.30–1.64 |
| 16q12.1  | Replication | 0.27        | 1,540, 1,046, 205        | 0.34     | 356, 372, 95         | 2.38 x 10^-6  | 1.62, 1.32–1.99 |
| TOX3     | Combined    | 0.26        | 2.797, 257, 181, 36      | 0.37     | 173, 204, 59         | 3.87 x 10^-15 | 1.50, 1.35–1.66 |

dis-equilibrium block of approximately 52 kb in length (Supplementary Fig. 4). RAD51 family members function in both mitotic and meiotic homologous recombination and in DNA double-strand break repair. rs999737, located in intron 10 of RAD51B, has previously been shown to be associated with risk of female breast cancer. This SNP maps approximately 335 kb telomeric to rs1314913 and is separated from it by strong recombination hotspots (Fig. 1a and Supplementary Fig. 4). rs999737 and rs1314913 were only weakly correlated in the male breast cancer cases ($r^2 = 0.02$) and the HapMap Utah residents of Northern and Western European ancestry (CEU) population ($r^2 = 0.006$). To formally test for independence of the rs1314913 association signal from that of rs999737, we fitted a logistic regression model with data from the discovery phase samples adjusted for rs999737, obtaining an OR for rs1314913 of 1.54 ($P = 1.04 	imes 10^{-9}$). Conversely, the OR for rs999737 after adjustment for rs1314913 was 0.93 ($P = 0.25$).

To provide further insight into the association at 14q24.1, we imputed genotypes in cases and controls using data from the 1000 Genomes Project. Fifty-two imputed SNPs were more strongly associated with male breast cancer than rs1314913 and delineated an 85-kb cluster from 67.68 Mb to 67.77 Mb (Fig. 1a and Supplementary Table 5). To examine whether any directly genotyped or imputed SNPs annotated a putative transcription factor–binding site or enhancer element, we conducted a bioinformatic search of the region (Supplementary Methods). Seven associated SNPs, including rs1314913, were highly evolutionarily conserved (Supplementary Table 6). Analysis of Encyclopædia of DNA Elements (ENCODE) project data, including the Broad histone modification data sets for human mammary epithelial cells (HMECs), showed that two conserved SNPs, rs1314913 and an adjacent SNP, rs1316014, were located in a transcription factor–binding site, lying within a DNase hypersensitive site flanked by regions of high histone H3 lysine 4 (H3K4) mono- or dimethylation and low trimethylation, features that are characteristic of enhancer elements (Supplementary Figs. 5 and 6). In silico predictions are compatible with the minor alleles of both rs1314913 and rs1316014 abrogating the DNA binding sites of activator protein (AP)-1 and related transcription factors (Supplementary Fig. 7). It is possible that the role of AP-1 in modulating estrogen signaling and transcription might explain the association between rs1314913 and male breast cancer.

We have previously shown in a much smaller study that rs3803662, a synonymous SNP in LOC643174 mapping to chromosome 16q12.1 at 51,143,812 bp, was associated with male breast cancer risk, albeit not at genome-wide levels of significance. Here, we provide robust confirmatory evidence of this association (Table 1). Examination of imputed data suggests that the association spans a 61-kb region from 51.09 Mb to 51.16 Mb that is proximal to LOC643174 (Fig. 1b), with the associated region localizing to the TOX3 gene (encoding TOX high-mobility-group family member 3).

Rare variants in two breast cancer susceptibility genes, BRCA2 and CHEK2, have larger ORs in males compared with females, and we show here that this is also true for two common susceptibility alleles. Both rs1314913 and rs3803662 are notable because of the magnitude of their effects. Comparing the breast cancer OR for rs3803662 in our data with the published estimate for females (OR = 1.20 (1.16–1.24))

Variants at 24 loci have so far been shown to influence female breast cancer risk.

Both rs1314913 and rs3803662 are notable because of the magnitude of their effects. Comparing the breast cancer OR for rs3803662 in our data with the published estimate for females (OR = 1.20 (1.16–1.24)), we found that the effect was significantly greater in males ($P = 7.76 	imes 10^{-5}$). Because rs1314913 represents a new breast cancer susceptibility locus, there are no equivalent estimates in females for comparison.

Variants at 24 loci have so far been shown to influence female breast cancer risk.

The associations of these variants with male breast cancer are shown in Supplementary Table 7. In addition to rs3803662, SNPs at 2q35, 6q25.1, 10q21.2, 11q13.3 and 12p11.22 were significantly associated with $P < 0.05$. Loci at 3p24.1, 9p21.3 and 14q24.1 showed borderline associations at $P \leq 0.1$. There was no significant association, however, between variants at the FGFR2 locus on chromosome 10q26.13 and male breast cancer risk (rs2981582; OR = 1.07 (0.96–1.20); $P = 0.21$). This observation is unexpected because male breast cancer is almost entirely estrogen receptor positive. rs2981582 is the SNP with the strongest known association with estrogen receptor–positive breast cancer in females, and the power of our study to detect an allele
with the same effect size as in female breast cancer at $P \leq 0.05$ was close to 100%. rs3803662, however, is strongly associated with both estrogen receptor–negative and estrogen receptor–positive breast cancer in females. Therefore, the estrogen receptor status of male breast cancers does not provide an obvious explanation for the observed SNP associations.

These data provide evidence for low-penetration susceptibility to male breast cancer. Given the modest size of our study, it is likely that additional risk variants can be identified by future GWAS.

**URLs.** Genotype Libraries and Utilities (GLU), http://code.google.com/p/glu-genetics/; R, http://cran.r-project.org/; UCSC Genome Browser, http://genome.ucsc.edu/; 1000 Genomes Project, http://www.1000genomes.org/; HapMap, http://hapmap.ncbi.nlm.nih.gov; SNAP, http://www.broadinstitute.org/mpg/snap/; IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html; SNPTESv2, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html; SHAPEIT, http://www.shapeit.fr; Haploview, http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview; EIGENSOFT v3, http://genepath.med.harvard.edu/~reich/Software.htm; TRANSFACT Matrix Database, http://www.gene-regulation.com/pub/databases.html; 1958 British Birth Cohort, http://www2.le.ac.uk/projects/birthcohort, http://www.bristol.ac.uk/alspac/, http://www.cls.ioe.ac.uk/ncds and http://www.esds.ac.uk/findingData/ncds.asp.

**Note:** Supplementary information is available in the online version of the paper.

**ACKNOWLEDGMENTS**

We thank the men who participated in the study. For ICR sample collection, we are grateful to B. Smith, D. Hogben, R. McCann, J. Melia, S. Squires, M. Snigorska, J. Micic and K. Sibley for administrative help and advice; to the research nurses A. Butlin, M. Pelerin, J. Wood and T. Gardener for collection of blood samples and questionnaire data from participants; to D. Thomas and her team from the Breakthrough Generations Study for coordinating collection of controls; and to the cancer registries of England and Wales for providing information on eligible participants. The Study of Epidemiology and Risk Factors in Cancer Heredity (SEARCH) thanks the SEARCH team. We wish to thank H. Thorne, E. Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Department of Pathology at the Westmead Millennium Institute of Cancer Research and Prevention Institute; L.O. and G.G. coordinated sample collection for the Finnish Male Breast Cancer Study; A.H.T. and P.A.J. coordinated sample collection at the Leeds Institute of Molecular Medicine; D.P., G.M. and I.Z. coordinated sample collection at ISPO (Cancer Research and Prevention Institute); L.O. and G.G. coordinated sample collection at Sapienza University of Rome; A.H. and A.M.W. coordinated sample collection at Erasmus University Medical Center; S.N. and M.K. coordinated sample collection at the Institute of Oncology Ljubljana; M.G.-D. and J.E.C. coordinated sample collection in Galicia, Spain; H.O. and I.H. coordinated sample collection at Lund University; D.F.E., P.D.P.P. and A.M.D. coordinated sample collection for the SEARCH/Cambridge studies; D.T.B. coordinated sample collection at the Leeds Institute of Molecular Medicine; and S.L.N. and L.S. coordinated collection of US samples. All authors contributed to the final manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2417.

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

1. Bevier, M., Sundquist, K. & Hemminki, K. Breast Cancer Res. Treat. 132, 723–728 (2012).
2. Basham, V.M. et al. Breast Cancer Res. 4, R2 (2002).
3. Thomas, G. et al. Nat. Genet. 41, 579–584 (2009).
4. Wang, C. et al. Mol. Endocrinol. 25, 1527–1538 (2011).
5. Orr, N. et al. PLoS Genet. 7, e1002290 (2011).
6. Meijers-Heijboer, H. et al. Nat. Genet. 31, 55–59 (2002).
7. Easton, D.F. et al. Nature 447, 1087–1093 (2007).
8. Haiman, C.A. et al. Nat. Genet. 43, 1210–1214 (2011).
9. Stevens, K.N. et al. Cancer Res. 72, 1795–1803 (2012).
10. Fletcher, C. et al. J. Natl. Cancer Inst. 103, 425–435 (2011).
11. Stacey, S.N. et al. Nat. Genet. 39, 865–869 (2007).
12. Stacey, S.N. et al. Nat. Genet. 40, 703–706 (2008).
13. Turnbull, C. et al. Nat. Genet. 42, 504–507 (2010).
14. Ghousaini, M. et al. Nat. Genet. 44, 312–318 (2012).
15. Garcia-Closas, M. et al. PLoS Genet. 4, e1000594 (2008).