Published Online First on March 24, 2009 as 10.1158/0008-5472.CAN-08-4266

Priority Report

A Breast Cancer Risk Haplotype in the Caspase-8 Gene

Neil Duncan Shephard,[1] Ryan Abo,[3] Sushila Harkisandas Rigas,[1] Bernd Frank,[4,5] Wei-Yu Lin,[1] Ian Wallace Brock,[1] Adam Shippen,[1] Sabapathy Prakash Balasubramanian,[2] Malcolm Walter Ronald Reed,[2] Claus Rainer Bartram,[6] Alfons Meindl,[8] Rita Katharina Schmutzler,[9] Christoph Engel,[10] Barbara Burwinkel,[4,7] Lisa Anne Cannon-Albright,[3] Kristina Allen-Brady,[3] Nicola Jane Camp,[3] and Angela Cox[1]

[1]Institute for Cancer Studies and [2]Academic Unit of Surgical Oncology, School of Medicine and Biomedical Sciences, University of Sheffield, Sheffield, United Kingdom; [3]Genetic Epidemiology, Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, Utah; Divisions of [4]Molecular Epidemiology and [5]Clinical Epidemiology and Aging Research, German Cancer Research Center; [6]Institute of Human Genetics and [7]Division Molecular Biology of Breast Cancer, Department of Gynaecology and Obstetrics, University of Heidelberg, Heidelberg, Germany; [8]Department of Gynaecology and Obstetrics, Klinikum rechts der Isar, Technische Universität, Munich, Germany; [9]Center of Familial Breast and Ovarian Cancer, Department of Obstetrics/Gynecology and Center of Integrated Oncology, University of Cologne, Cologne, Germany; and [10]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany

Abstract

Recent large-scale studies have been successful in identifying common, low-penetrance variants associated with common cancers. One such variant in the caspase-8 (*CASP8*) gene, *D302H* (rs1045485), has been confirmed to be associated with breast cancer risk, although the functional effect of this polymorphism (if any) is not yet clear. In order to further map the *CASP8* gene with respect to breast cancer susceptibility, we performed extensive haplotype analyses using single nucleotide polymorphisms (SNP) chosen to tag all common variations in the gene (tSNP). We used a staged study design based on 3,200 breast cancer and 3,324 control subjects from the United Kingdom, Utah, and Germany. Using a haplotype-mining algorithm in the UK cohort, we identified a four-SNP haplotype that was significantly associated with breast cancer and that was superior to any other single or multi-locus combination ($P = 8.0 \times 10^{-5}$), with a per allele odds ratio and 95% confidence interval of 1.30 (1.12–1.49). The result remained significant after adjustment for the multiple testing inherent in mining techniques (false discovery rate, $q = 0.044$). As expected, this haplotype includes the *D302H* locus. Multicenter analyses on a subset of the tSNPs yielded consistent results. This risk haplotype is likely to carry one or more underlying breast cancer susceptibility alleles, making it an excellent candidate for resequencing in homozygous individuals. An understanding of the mode of action of these alleles will aid risk assessment and may lead to the identification of novel treatment targets in breast cancer. [Cancer Res 2009;69(7):2724–8]

Introduction

Recent genome-wide and candidate gene association studies have started to convincingly identify low-penetrance variants associated with breast cancer (1–4). The only confirmed common variant that has emerged from candidate gene studies for breast cancer thus far is in the gene for the apoptosis-related cysteine protease caspase-8 (*CASP8*), located on chromosome region 2q33 (2, 5, 6). The rare allele of the nonsynonymous variant *D302H* (rs1045485) was associated with a reduced risk of breast cancer, with a per allele odds ratio (OR) and 95% confidence interval (95% CI) of 0.88 (0.84–0.92) in a large study of 16,423 cases and 17,109 controls carried out by the Breast Cancer Association Consortium (2). As yet, there is no known functional effect of rs1045485, and it is nonpolymorphic in Asian populations. Another *CASP8* polymorphism, a 6-bp insertion-deletion (indel) in the promoter of *CASP8* (rs3834129) was found to reduce breast cancer risk in a Chinese population (7). However, subsequent larger studies failed to replicate this finding (8, 9).

The aim of the present work was to use a single nucleotide polymorphism (SNP)–tagging approach to further map the *CASP8* gene with respect to breast cancer risk in order to move towards the identification of potential susceptibility variant(s) (10).

Materials and Methods

Case and control subjects. The primary set of case and control subjects were drawn from the Sheffield Breast Cancer Study (SBCS) and consisted of histopathologically confirmed breast cancer patients recruited from the surgical outpatient clinics of the Royal Hallamshire Hospital, Sheffield, United Kingdom between November 1998 and January 2005. Controls were recruited from patients attending the Sheffield Mammography Screening Service between September 2000 and January 2004, whose mammograms showed no evidence of breast lesions. All cases and controls were of North European origin and resident in the Sheffield area (5, 11). The second set comprised unrelated *BRCA1/2* mutation–negative breast cancer patients recruited between 1997 and 2007 by three centers from the German Consortium for Hereditary Breast and Ovarian Cancer (GC-HBOC; refs. 6, 8). All patients had been screened for mutations in the *BRCA1* and *BRCA2* genes by denaturing high-performance liquid chromatography analysis of all exons followed by direct sequencing. Ethnically matched controls were selected from unrelated healthy female blood donors collected by the Institute of Transfusion Medicine and Immunology (Mannheim, Germany) between the years 2004 and 2007. The Utah Breast Cancer Study (UBCS) cohort consisted of *BRCA1/2* mutation–negative cases (established by sequencing, family inference, or linkage evidence) from extended high-risk Utah pedigrees ascertained using the Utah Population Database (12). Controls were unrelated birth cohort– and sex-matched cancer-free individuals.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

N.D. Shephard, R. Abo, N.J. Camp, and A. Cox contributed equally to this work.

Requests for reprints: Angela Cox, Institute for Cancer Studies, School of Medicine and Biomedical Sciences, University of Sheffield, Beech Hill Road, Sheffield S10 2RX, United Kingdom. Phone: 44-114-271-2373; Fax: 44-114-271-3892; E-mail: a.cox@shef.ac.uk.

doi:10.1158/0008-5472.CAN-08-4266

Table 1. Association statistics for SNPs in the *CASP8* gene

		SBCS*				Meta-analysis of three studies*			
SNP [†]	Position (bp) [‡]	MAF[§]	OR (95% CI) heterozygotes	OR (95% CI) rare homozygotes	P_{trend}	OR (95% CI) heterozygotes	OR (95% CI) rare homozygotes	P_{trend}	P_{hom}[‖]
rs3834129	201805777	0.532	0.90 (0.73–1.11)	0.77 (0.61–0.97)	0.027	0.95 (0.83–1.08)	0.82 (0.70–0.95)	0.008	0.33
rs3769826	201811294	0.450	1.05 (0.87–1.28)	1.23 (0.97–1.55)	0.099				
rs7608692	201819204	0.214	0.94 (0.79–1.12)	1.05 (0.68–1.61)	0.68				
rs3820972	201819265	0.085	1.11 (0.86–1.44)	0.80 (0.33–1.95)	0.64				
rs3769825	201819625	0.420	1.10 (0.91–1.32)	1.26 (0.99–1.60)	0.063				
rs13402616	201828012	0.062	1.12 (0.87–1.44)	1.45 (0.46–4.57)	0.30				
rs1861269	201835183	0.036	0.87 (0.63–1.21)	—	0.52				
rs6435074	201836192	0.245	1.22 (1.02–1.45)	1.17 (0.84–1.64)	0.046	1.12 (1.00–1.27)	1.22 (0.97–1.55)	0.013	0.93
rs6723097	201836863	0.351	1.15 (0.96–1.37)	1.36 (1.04–1.76)	0.017	1.14 (1.00–1.30)	1.34 (1.11–1.61)	0.0008	0.99
rs3754934	201840352	0.048	0.90 (0.66–1.23)	3.06 (0.32–29.44)	0.73				
rs3817578	201844840	0.049	1.03 (0.77–1.38)	1.06 (0.07–16.94)	0.86				
rs1045485	201857834	0.156	0.87 (0.72–1.05)	0.59 (0.32–1.09)	0.041	0.91 (0.79–1.05)	0.65 (0.41–1.04)	0.043	0.48
rs1045487	201857941	0.043	0.96 (0.71–1.30)	3.04 (0.32–29.31)	0.97				
rs13113	201860407	0.450	0.94 (0.78–1.14)	0.90 (0.71–1.15)	0.39				

*OR and 95% CIs are relative to the common homozygous reference genotype.
†For rs3834129, ins/ins was used as the reference genotype to be consistent with published data (8). Largely overlapping data for rs3834129 and rs1045485 in SBCS and GC-HBOC have been published previously (2, 6, 8).
‡Positions are derived from dbSNP build 126.
§Minor allele frequency is given for SBCS controls.
‖Test for homogeneity between studies.

Selection of tSNPs. All available HapMap[11] SNPs within a 50 kb region spanning the *CASP8* gene, and SNPs from dbSNP[12] with a minor allele frequency of >0.05, were genotyped on 135 random SBCS control samples. The optimal set of 12 tSNPs was identified from these data by principal components analysis (13). The 12 tSNPs were supplemented by two further SNPs identified by DNA sequencing of regions containing putative SNPs, plus the 6 bp promoter indel variant rs3834129. Thus, a total of 15 SNPs were selected for genotyping.

Genotyping. Genotyping was carried out using the Applied Biosystems SNPlex multiplex system (SBCS samples) or 5′ nuclease PCR (UBCS and GC-HBOC samples). The 6 bp indel was genotyped by fragment analysis on an ABI 3730 automated sequencer. Genotyping quality was assessed by examination of duplicate concordance and call rates for each SNP and a test for compliance with Hardy-Weinberg equilibrium (HWE) in controls. A summary of genotyping quality data is shown in Supplementary Table S1. SNPs with duplicate concordance rates of <98%, call rates <90%, or P_{HWE} <0.005 were removed from the analysis.

Statistical analysis. All statistical tests were two-sided. Evidence of association for single SNPs in the primary discovery set was initially assessed by use of a trend test. Per allele and genotypic OR and 95% CIs were estimated within a logistic regression framework with the common homozygotes as reference group. In order to account for familial relatedness in the UBCS subjects, meta-analyses of individual SNPs across study populations were carried out using the Genie software package which uses Monte Carlo testing to derive empirical estimates of significance and CIs (14, 15).

Pairwise R^2 and D' values were estimated based on genotype data from 123 SBCS controls using Haploview (16). Haplotype frequencies were estimated by use of the estimation maximization algorithm within SNPHAP.[13] The hapConstructor module of Genie was used to build

[11] http://www.hapmap.org/
[12] http://www.ncbi.nlm.nih.gov/
[13] http://www-gene.cimr.cam.ac.uk/clayton/software/

Table 2. Levels of correlation between breast cancer–associated SNPs

SNP*	rs3834129	rs6435074	rs6723097	rs1045485
rs3834129	—	0.20	0.35	0.06
rs6435074	0.69	—	0.67	0.05
rs6723097	0.73	1.00	—	0.08
rs1045485	0.66 [†]	1.00 [†]	1.00 [†]	—

*Pairwise R^2 values (*top right*), and pairwise D' values (*bottom left*).
†LOD < 2.0.

Table 3. Haplotype frequencies for SBCS

Haplotype	rs3834129*	rs3769826	rs7608692	rs3820972	rs3769825	rs13402616	rs1861269	rs6435074*
1	2	1	1	1	1	1	1	1
2†	1	2	1	1	2	1	1	2
3	2	1	1	1	1	1	1	1
4	2	1	1	1	1	1	1	1
5	1	2	2	1	2	1	1	2
6	1	2	2	1	1	1	1	1
7	1	2	2	1	2	1	1	1
8†	1	1	1	1	2	2	1	1
9	1	2	1	2	2	1	1	1
10	1	2	2	1	2	1	1	1
11†	1	2	1	1	1	1	1	2
12	1	2	1	2	2	1	1	1
13	1	2	2	1	2	2	2	1
14†	1	2	1	2	2	1	1	2
15	1	2	1	2	2	1	1	1

*Available in three-study meta-analysis.
†1-1-2-1 risk haplotype.

combinations of SNPs associated with breast cancer (17). This data-mining module includes tests for dominant, additive, recessive, and allelic models for each haplotype with OR, χ^2 and χ^2_{trend} statistics calculated. Individuals with >50% missing genotype data were excluded from the analysis. In the remaining individuals, missing genotypes were internally imputed, and the haplotypes were estimated via the estimation maximization algorithm. The significance thresholds used for the haplotype construction process were 0.05, 0.005, 0.0005, 0.0001 for haplotypes of one to four markers, respectively, and 0.00005 thereafter. Construction-wide false discovery rate (FDR) q values for the best haplotypes, that appropriately account for the construction process, were determined empirically using 100,000 simulations.

Polytomous logistic regression and logistic regression (stratified by study) were used to compare genotype frequencies in different subgroups of cases, based on an additive model for genotype as above. Likelihood ratio testing was used to compare models with and without terms for genotype.

Results

We applied a staged study design based on three case-control population sets; the primary, discovery set (SBCS), and two additional sets to establish the robustness of findings (GC-HBOC and UBCS). A total of 14 SNPs were successfully genotyped in 1,228 case and 1,222 control subjects in the SBCS discovery set (Supplementary Tables S1 and S2). Four SNPs (rs3834129, rs6435074, rs6723097, and rs1045485) showed significant associations with breast cancer ($P_{trend} < 0.05$), with rs6723097 being the most significant, with per allele OR (95% CI) of 1.16 (1.03–1.31; $P_{trend} = 0.017$; Table 1). These four SNPs were genotyped in samples from 1,220 cases and 1,664 controls from GC-HBOC and 752 cases and 438 controls from UBCS (Supplementary Table S2). Three of the four SNPs yielded smaller empirical P_{trend} values in the three-study meta-analysis compared with SBCS alone, with rs6723097 again yielding the most significant result ($P_{trend} = 0.0008$), with no evidence of heterogeneity between studies (Table 1). Table 2 shows that there is generally a low degree of pairwise correlation between the four SNPs, with the exception of rs6723097 and rs6435074 ($R^2 = 0.67$). As expected, the D' values are somewhat higher, suggesting that the associated SNPs may be marking one or more underlying breast cancer haplotypes.

With the aim of identifying any such haplotypes that might carry functional etiological variants, we searched for susceptibility haplotypes using the hapConstructor module of Genie in the SBCS data set (17). Table 3 shows a summary of all haplotypes with a frequency of >1% in SBCS. HapConstructor identified a four-locus haplotype 1-1-2-1 at rs7608692, rs1861269, rs6723097, and rs3817578 as being most significant ($P = 8.0 \times 10^{-5}$), with a per allele OR (95% CI) of 1.30 (1.12–1.49), and a construction-wide FDR q value of 0.044. This four-allele haplotype has a frequency of 19.8% in controls and 24.2% in cases and is present on haplotypes 2, 8, 11, and 14 (Table 3). The only other four-locus haplotype to surpass the significance thresholds set in the data-mining process was identical at the first three SNP positions and replaced rs3817578 with *D302H* (rs1045485; $P = 1.0 \times 10^{-4}$). These two haplotypes constituted 16 of the 18 significant tests that were contained in the group of tests with an FDR of 0.044. Hence, there is extremely good evidence that these related haplotypes are true indicators of an underlying susceptibility variant. Furthermore, in a stepwise logistic regression, the 1-1-2-1 haplotype alone provided the best fitting model, compared with models involving any of the individual SNPs.

To assess the robustness of these results, we also carried out a meta-haplotype construction with the four SNPs typed in the three study populations (rs3834129, rs6435074, rs6723097, and rs1045485). HapConstructor extracted a two-SNP haplotype across rs6723097 and rs1045485 (1, 2) as the most significant ($P = 2.0 \times 10^{-5}$; FDR q value 0.002), with a protective per allele OR (95% CI) of 0.76 (0.68–0.85). The complement of this haplotype, 2-1, increased risk and was also significant (3.3×10^{-4}), with a per allele OR (95% CI) of 1.15 (1.06–1.24). This two-SNP combination also lies on haplotypes 2, 8, 11, and 14 (Table 3), and these two SNPs are also found on the four-locus risk haplotype in the

Table 3. Haplotype frequencies for SBCS (Cont'd)

rs6723097*	rs3754934	rs3817578	rs1045485*	rs1045487	rs13113	Controls	Cases
1	1	1	1	1	2	0.2922	0.2897
2	1	1	1	1	1	0.1163	0.1396
1	1	1	2	1	1	0.1123	0.0931
1	1	1	1	1	1	0.0696	0.0668
2	1	1	1	1	1	0.0664	0.0585
1	1	1	1	1	2	0.0456	0.0430
1	1	1	1	1	2	0.0321	0.0335
2	1	1	1	1	1	0.0243	0.0322
1	1	1	1	1	2	0.0275	0.0257
2	2	2	1	2	1	0.0203	0.0233
2	1	1	1	1	1	0.0195	0.0208
1	1	1	2	1	1	0.0193	0.0212
2	1	1	1	1	2	0.0192	0.0174
2	1	1	1	1	1	0.0153	0.0181
2	2	2	1	2	1	0.0175	0.0122

discovery analysis. Thus, the meta-analysis haplotypic associations are extremely consistent with the four-allele haplotype association seen in SBCS.

A case-only meta-analysis across the three studies yielded no evidence that either the individual SNPs or the haplotypes were associated with age at onset, family history, bilateral disease, or estrogen or progesterone receptor tumor status (data not shown).

Discussion

Our haplotype-mining results, based on three independent data sets, provide evidence that an extended multilocus *CASP8* haplotype is associated with breast cancer. The risk haplotype provides a better fitting model than any combination of the individual SNPs. This suggests that additional untyped variants carried on this haplotype may be responsible for the increased breast cancer risk. Resequencing of DNA samples from individuals carrying the high-risk and low-risk haplotypes should allow the underlying causative variants to be identified. Such variants might affect the molecular interactions of caspase-8, caspase-8 activity (coding variants), or caspase-8 levels via effects on transcription factor binding, RNA splicing, or RNA stability (intronic/intergenic variants).

Aside from a well-defined role as an initiator of apoptosis, caspase-8 has been proposed as a molecular switch between cell motility (promoted by procaspase-8) and apoptosis (promoted by mature caspase-8) (18). Caspase-8 processing to the mature form is in turn controlled by phosphorylation by c-SRC, a proto-oncogene tyrosine kinase whose activity is up-regulated in many types of tumor (19). It will be important to determine whether cancer-associated variants in *CASP8* affect these processes. Furthermore, it is intriguing to note that although the rare allele of *CASP8 D302H* is associated with a decreased risk of breast cancer, it is associated with an increased risk of glioma (20). Further studies including more comprehensive SNP panels and cancer characteristics are therefore needed to help us understand the roles of caspase-8 in different cancer types.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 11/7/08; revised 1/30/09; accepted 2/2/09; published OnlineFirst 3/24/09.

Grant support: Genotyping and data analysis in Sheffield, UK were supported by the Breast Cancer Campaign (grant no. 2004Nov49) and Yorkshire Cancer Research (grant no. S295). For the Utah Breast Cancer Study, genotype data and analysis were supported by a Susan G. Komen Foundation grant (BCTR0706911) and an NIH grant (CA98364). Recruitment in Utah was supported in part by the Utah Cancer Registry and the Utah Population Database. The Utah Cancer Registry is funded by contract N01-PC-35141 from the National Cancer Institute's SEER program with additional support from the Utah State Department of Health and the University of Utah. Partial support for the Utah Population Database was provided by the University of Utah Huntsman Cancer Institute.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank all study subjects for their participation in this research; Helen Cramp and Dan Connley for subject recruitment and data collection in Sheffield; and Sandrine Tchatchou for overseeing the DNA sample collection for GC-HBOC.

References

1. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–93.
2. Cox A, Dunning AM, Garcia-Closas M, et al. A common coding variant in *CASP8* is associated with breast cancer risk. Nat Genet 2007;39:352–8.
3. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007;39:865–9.
4. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–4.
5. MacPherson G, Healey CS, Teare D, et al. Association of a common variant of the *CASP8* gene with reduced risk of breast cancer. J Natl Cancer Inst 2004;96:1866–9.
6. Frank B, Bermejo JL, Klaes R, et al. Re: Association of a common variant of the *CASP8* gene with reduced risk of breast cancer. J Natl Cancer Inst 2005;97:1012.
7. Sun T, Gao Y, Tan W, et al. A six-nucleotide insertion-deletion polymorphism in the *CASP8* promoter is

associated with susceptibility to multiple cancers. Nat Genet 2007;39:605.
8. Frank B, Rigas SH, Bermejo JL, et al. The *CASP8 −652 6N del* promoter polymorphism and breast cancer risk: a multicenter study. Breast Cancer Res Treat 2008;111: 139–44.
9. Haiman CA, Garcia RR, Kolonel LN, Henderson BE, Wu AH, Le Marchand L. A promoter polymorphism in the *CASP8* gene is not associated with cancer risk. Nat Genet 2008;40:259.
10. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217–23.
11. Rafii S, O'Regan P, Xinarianos G, et al. A potential role for the *XRCC2 R188H* polymorphic site in DNA-damage repair and breast cancer. Hum Mol Genet 2002;12: 1433–8.
12. Cannon-Albright LA, Thomas A, Goldgar DE, et al. Familiality of cancer in Utah. Cancer Res 1994;54: 2378–85.
13. Horne BD, Camp NJ. Principal component analysis for selection of optimal SNP-sets that capture intragenic genetic variation. Genet Epidemiol 2004;26:11–21.
14. Allen-Brady K, Wong J, Camp NJ. PedGenie: an analysis approach for genetic association testing in extended pedigrees and genealogies of arbitrary size. BMC Bioinformatics 2006;7:209.
15. Curtin K, Wong J, Allen-Brady K, Camp NJ. PedGenie: meta genetic association testing in mixed family and case-control designs. BMC Bioinformatics 2007;8:448.
16. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–5.
17. Abo R, Knight S, Wong J, Cox A, Camp NJ. hapConstructor: automatic construction and testing of haplotypes in a Monte Carlo framework. Bioinformatics 2008:24;2105–7.
18. Frisch SM. Caspase-8: fly or die. Cancer Res 2008;68: 4491–3.
19. Cursi S, Rufini A, Stagni V, et al. Src kinase phosphorylates caspase-8 on Tyr380: a novel mechanism of apoptosis suppression. EMBO J 2006;25:1895–905.
20. Bethke L, Sullivan K, Webb E, et al. The common *D302H* variant of *CASP8* is associated with risk of glioma. Cancer Epidemiol Biomarkers Prev 2008;17:987–9.