

Fine mapping of 14q24.1 breast cancer susceptibility locus

Phoebe Lee · Yi-Ping Fu · Jonine D. Figueroa · Ludmila Prokunina-Olsson · Jesus Gonzalez-Bosquet · Peter Kraft · Zhaoming Wang · Kevin B. Jacobs · Meredith Yeager · Marie-Josèphe Horner · Susan E. Hankinson · Amy Hutchinson · Nilanjan Chatterjee · Montserrat Garcia-Closas · Regina G. Ziegler · Christine D. Berg · Sandra S. Buys · Catherine A. McCarty · Heather Spencer Feigelson · Michael J. Thun · Ryan Diver · Ross Prentice · Rebecca Jackson · Charles Kooperberg · Rowan Chlebowski · Jolanta Lissowska · Beata Peplonska · Louise A. Brinton · Margaret Tucker · Joseph F. Fraumeni Jr. · Robert N. Hoover · Gilles Thomas · David J. Hunter · Stephen J. Chanock

Received: 14 June 2011 / Accepted: 2 September 2011 / Published online: 30 September 2011
© Springer-Verlag (outside the USA) 2011

Abstract In the National Cancer Institute Cancer Genetic Markers of Susceptibility (CGEMS) genome-wide association study of breast cancer, a single nucleotide polymorphism (SNP) marker, rs999737, in the 14q24.1 interval, was associated with breast cancer risk. In order to fine map this region, we imputed a 3.93 MB region flanking rs999737 for Stages 1 and 2 of the CGEMS study (5,692 cases, 5,576 controls) using the combined reference panels of the HapMap 3 and the 1000 Genomes Project. Single-marker association testing and variable-sized sliding-window haplotype analysis were performed, and for both analyses the initial tagging SNP rs999737 retained the strongest association with breast cancer risk. Investigation

of contiguous regions did not reveal evidence for an additional independent signal. Therefore, we conclude that rs999737 is an optimal tag SNP for common variants in the 14q24.1 region and thus narrow the candidate variants that should be investigated in follow-up laboratory evaluation.

Introduction

Breast cancer is the most commonly diagnosed cancer in women after non-melanoma skin cancer, and is the second leading cause of cancer deaths among women in the US (Jemal et al. 2010). The contribution of genetic susceptibility to breast cancer has been the subject of intense study for decades. Initially, highly penetrant mutations in *BRCA1* and *BRCA2* were identified in families with multiple affected individuals, but together these account for less than 5% of breast cancer overall in the US (Hall et al. 1990; Pruthi et al. 2010). The risk of a first-degree relative with

P. Lee and Y.-P. Fu contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-011-1088-4) contains supplementary material, which is available to authorized users.

P. Lee · Y.-P. Fu · J. D. Figueroa · L. Prokunina-Olsson · Z. Wang · K. B. Jacobs · M. Yeager · M.-J. Horner · A. Hutchinson · N. Chatterjee · M. Garcia-Closas · R. G. Ziegler · L. A. Brinton · M. Tucker · J. F. Fraumeni Jr. · R. N. Hoover · S. J. Chanock (✉)
Division of Cancer Epidemiology and Genetics, Department of Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA
e-mail: chanocks@mail.nih.gov

J. Gonzalez-Bosquet
Gynecologic Oncology, Department of Women's Oncology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA

P. Kraft · D. J. Hunter
Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA

Z. Wang · K. B. Jacobs · M. Yeager · A. Hutchinson
Core Genotyping Facility, Advanced Technology Program, SAIC-Frederick Inc., NCI Frederick, Frederick, MD 21701, USA

S. E. Hankinson · D. J. Hunter
Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

C. D. Berg
Division of Cancer Prevention, NCI, NIH, DHHS, Bethesda, MD 20892, USA

S. S. Buys
Department of Internal Medicine, University of Utah, Salt Lake City, UT 84132, USA

breast cancer approximately doubles compared to women with no known family history of breast cancer, suggesting a more complex genetic architecture could contribute to breast cancer risk in the general population (Lichtenstein et al. 2000; Pruthi et al. 2010).

In the last 4 years, genome-wide association studies (GWAS) have successfully discovered at least 20 distinct loci associated with risk for breast cancer (Ghoussaini and Pharoah 2009; Varghese and Easton 2010) and additional regions are expected to be discovered as meta-analyses proceed (Park et al. 2010). Since each region identified in breast cancer GWAS appears to contribute a small fraction to the overall risk, it is not surprising that the current set explains only a small fraction of the overall familial risk for breast cancer (Ghoussaini and Pharoah 2009). Nonetheless, the discoveries in GWAS point towards new biological insights into the underlying factors contributing to a complex disease, such as breast cancer. The follow-up for each region conclusively identified by GWAS requires fine mapping of common variants not directly tested in the GWAS to determine if one or more additional variants could be directly responsible for the signal. Since linkage disequilibrium (LD) patterns for common variants can often reveal multiple surrogates for any given marker, fine-mapping is necessary to nominate the subset of variants of high interest for functional investigation of the underlying biological mechanisms responsible for the direct association with breast cancer.

The Cancer Genetic Markers of Susceptibility (CGEMS) GWAS reported a single nucleotide polymorphism (SNP), rs999737, within the 14q24.1 interval that maps to an intron of the *RAD51*-like 1 (*RAD51L1*) gene. Since the most significant tagging signal rs999737 localized to an intron in *RAD51L1*, we conducted a fine-mapping analysis of the 14q21 region in order to better define the chromosomal

boundaries of interest for subsequent biological experiments designed to explain the association signal. *RAD51L1* is one of five paralogs to the gene *RAD51*, which encodes the eukaryotic DNA recombination repair protein (Thacker and Zdzienicka 2004; Thomas et al. 2009). An independent pooled analysis in the Breast Cancer Association Consortium, using data from over 40,000 cases and 40,000 controls, confirmed the association between the SNP rs999737 and breast cancer risk; the data also supported this SNP to be a risk factor for the major breast cancer subtypes, as defined by expression of estrogen-receptor, progesterone receptor, and Her2/neu (Figueroa et al. 2011). In contrast, this SNP did not modify risk for *BRCA1* or *BRCA2* mutation carriers (Antoniou et al. 2011). Molecular studies of the *RAD51L1* protein suggest its main function is to promote homologous recombination repair; specifically, the protein contributes to genome integrity maintenance in response to UV-induced damages (Havre et al. 1998; Takata et al. 2000). In breast cell lines, it has been shown that both transcription and protein levels of *RAD51L1* are down-regulated by over-expression of *EZH2*, a transcriptional repressor previously linked to aggressive and metastatic breast cancers; *RAD51L1* may also partner with the protein *EVL* (Ena/Vasp-like) to stimulate inappropriate *RAD51*-mediated recombination in breast cancer cell lines (Takaku et al. 2009; Zeidler et al. 2005).

To fine map the region surrounding rs999737 on 14q24.1, we used Stage 1 and 2 datasets from the CGEMS breast cancer GWAS to conduct an imputation analysis of SNP markers that were not directly genotyped in CGEMS. The imputation used 1000 Genomes Project (Durbin et al. 2010) and HapMap3 CEU (The International HapMap3 Consortium 2010) data. Here, we report the results of our work: the SNP marker rs999737 retains the strongest association of all imputed and genotyped SNPs across the 3.93 MB region and the data point toward a single signal on 14q24.1.

C. A. McCarty
The Center for Human Genetics, Marshfield Clinic Research
Foundation, Marshfield, WI 54449, USA

H. S. Feigelson
Institute for Health Research, Kaiser Permanente Colorado,
Denver, CO 80237, USA

M. J. Thun · R. Diver
Department of Epidemiology and Surveillance Research,
American Cancer Society, Atlanta, GA 30329, USA

R. Prentice · C. Kooperberg
Fred Hutchinson Cancer Research Center,
Seattle, WA 98109, USA

R. Jackson
Division of Diabetes, Endocrinology and Metabolism,
The Ohio State University Medical Center, Columbus,
OH 43210, USA

R. Chlebowski
Harbor-University of California at Los Angeles Medical Center,
Torrance, CA 90509, USA

J. Lissowska
Department of Cancer Epidemiology and Prevention,
M. Sklodowska-Curie Memorial, Cancer Center and Institute
of Oncology, Warsaw, Poland

B. Peplonska
Nofer Institute of Occupational Medicine, Lodz, Poland

G. Thomas
Foundation Synergie-Lyon-Cancer, U590 INSERM,
Centre Léon Bérard, 69373 Lyon, France

D. J. Hunter
Broad Institute of Harvard and MIT,
Cambridge, MA 02142, USA

Materials and methods

Study population

As described previously (Thomas et al. 2009), all study subjects analyzed in the CGEMS Breast Cancer Genome-Wide Association Scan were of European ancestry. Genotype data from Stage 1 and 2 of the CGEMS scan were used in this study. In Stage 1, 1,145 postmenopausal breast cancer cases and 1,142 healthy controls were genotyped with Human Hap500 Infinium Assay (Illumina) from the Nurses' Health Study (NHS) (Hunter et al. 2007). For Stage 2, a custom-designed iSelect assay from Illumina was used to genotype 8,981 subjects from 4 additional study groups: the Women's Health Initiative (WHI) (2,395 cases/2,410 controls); the Prostate, Lung, Colon and Ovarian Screening Trial (PLCO) Breast Cancer Study (948 cases/975 controls); the American Cancer Society Cancer Prevention Study II Nutrition Cohorts (CPSII) (535 cases/543 controls); and the Polish Breast Cancer Study (PBCS1) (669 cases/506 controls). For the combination of Stage 1 and Stage 2, a total of 5,692 breast cancer cases and 5,576 controls were available for further imputation and the subsequent association testing.

Imputation analysis

To infer potential genomic variants tagged by the SNP rs999737, we used the hidden Markov model program, IMPUTE2 (Howie et al. 2009), to impute genotypes based on HapMap Phase 3 CEU data (The International HapMap3 Consortium 2010) and 1000 Genomes Project (Durbin et al. 2010). Genotyped markers within a region of approximately 3.93 MB flanking rs999737 and *RAD51L1* (chr14q21: 66,173,322–70,099,944, UCSC genome build hg18) were available for analysis, including 770 SNPs from CGEMS Stage 1 data and 41 SNPs from Stage 2 in 11,268 individuals. Based on the combined reference panels of the 1000 Genomes Projects (2010 June release) and HapMap3 CEU data (2009 February release 2), an additional 11,047 putative SNPs were imputed for Stage 1 and 2 subjects across this region of 14q24.1. The average allele dosage was estimated with 100 iterations of the imputation algorithm conditional on a set of known haplotypes while simultaneously estimating the recombination map. The best-guessed genotypes (those with highest posterior probabilities) at each SNP for each subject were also generated using GTOOL (<http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html>). In addition to the average posterior probability for the best-guessed genotypes, the IMPUTE2-info score, which is associated with the imputed allele frequency estimate ranging from 1 (high confidence) to 0 (decreasing confidence), was also estimated for each SNP to evaluate imputation performance.

Statistical analysis

Fisher's exact tests for fitness for Hardy–Weinberg proportion (also known as Hardy–Weinberg Equilibrium, HWE) were evaluated across controls for all markers. Those that significantly deviated from HWE ($P < 10^{-6}$) were discarded from further analysis. LD measures (D' and r^2) and LD plots were estimated using Haploview (Barrett et al. 2005) for the control samples. Based on the per marker quality measures generated from IMPUTE2, markers with maximum posterior genotype probability < 0.9 were also filtered out to obtain reliable association results. The best-guessed SNP genotypes (those with highest posterior probabilities ≥ 0.9 , imputation certainty ≥ 0.9 , and did not significantly deviate from HWE) for all imputed markers were tested for association with breast cancer risk under an additive model using PLINK v1.07 (Purcell et al. 2007), adjusting for 10-year age groups, study sites, and 4 significant principal components to account for population heterogeneity as determined in the initial scan (Hunter et al. 2007; Thomas et al. 2009). A two-SNP model was used to determine whether an additional independent signal could be found by conditioning on the original CGEMS hit, rs999737. To explore the possible difference of the effect between different subtypes of breast cancer, we conducted a stratified analysis according to estrogen receptor (ER) and progesterone receptor (PR) status, under an additive genetic model adjusted for the already defined covariates. An analysis to account for imputation uncertainty using estimated allelic dosage was also performed with SNPTTEST v2 (<http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html>) and showed similar results (data not shown). A 2- to 8-SNP sliding-window haplotype analysis was performed across the region using PLINK (Purcell et al. 2007). Odds ratios and P values were estimated for each haplotype versus all others, adjusted for the effects of same covariates. SequenceLDhot (Fearhead 2006) was used to compute likelihood ratios (LR) for a set of putative recombination hotspots across the *RAD51L1* region, repeated with 5 non-overlapping sets of 100 controls and using background recombination rates estimated by PHASE v2.1 (Stephens et al. 2001). Snp.plotter (Luna and Nicodemus 2007), an R-based package, was used to simultaneously display P values from single-marker and haplotype analyses.

Results

Imputation and quality control

A total of 10,273 SNP markers within a 3.93 MB flanking region (chr14:66,173,332–70,099,944) of rs999737 were

imputed for 11,268 samples from the CGEMS Stages 1 and 2 of the breast cancer GWAS. Table 1 summarizes the genotyped and imputed SNPs across the 3.93 Mb region of *RAD51L1*. More than 60% of the imputed SNPs had a minor allele frequency greater than 1%. The overall IMPUTE2-info score was calculated to be 0.4347, the median value of all imputed markers using a measure of the performance of imputation with values ranging from 0 to 1 in ascending confidence. We did not remove SNPs with minor allele frequencies less than 5% in order to retain the less common variants that could provide evidence for a synthetic association in the region (Dickson et al. 2010). The association analysis was restricted to 464 SNPs, the subset of SNPs with an IMPUTE2-info score of ≥ 0.9 (a rigorous metric for high quality imputed genotype calls), and no evidence for deviation from HWE ($P > 10^{-6}$). Among these SNPs, 36 were directly genotyped in both Stage 1 and 2 samples, 31 were genotyped in Stage 1 and imputed in Stage 2 and 397 were imputed based on the 1000 Genome Project and HapMap 3 data. It is notable that a subset of the SNPs genotyped in Stage 1 could not be accurately imputed in Stage 2 using the public databases.

Association analysis of surrogates for rs999737

Figure 1a shows results from single-marker analysis of all imputed and genotyped markers within chromosomal region chr14: 67795579–68450029. Among them, 46 SNPs (including, the sentinel tagging SNP, rs999737) had IMPUTE2-info scores ≥ 0.9 and did not deviate from Hardy–Weinberg equilibrium; these were also nominally associated with breast cancer risks at $P_{\text{adj}} < 0.05$, yet none of the values approached genome-wide significance (Table 2). 11 of the significantly associated SNPs were genotyped, while the remaining 35 were inferred from IMPUTE2 using a combined HapMap3 and 1000 Genome reference panel. Within this group, the SNP with the strongest association with breast cancer risk was the initial GWAS SNP rs999737 ($P_{\text{adj}} = 8.23 \times 10^{-6}$; $\text{OR}_{\text{adj}} = 0.87$, 95% CI = 0.81–0.92), as reported for Stages 1 and 2, but not Stage 3, which conducted single SNP analysis (Thomas et al. 2009). We performed a sequential analysis for the other 45 SNPs conditioned on rs999737 ($P_{\text{adj}|\text{RS999737}}$); 32 SNPs remained nominally associated with breast cancer risk ($P_{\text{adj}|\text{RS999737}} < 0.05$; Table 3). 5 of the SNPs were genotyped in the original GWAS report, while the remaining 28 SNPs were inferred from IMPUTE2. To determine whether the list of 45 SNPs showing moderate association with breast cancer risk were independent signals from rs999737, we examined whether any SNPs showing significant association with breast cancer risks fell across the gene *RAD51L1*, or shared significant pair-wise LD with the initial GWAS SNP marker, rs999737. We discovered that

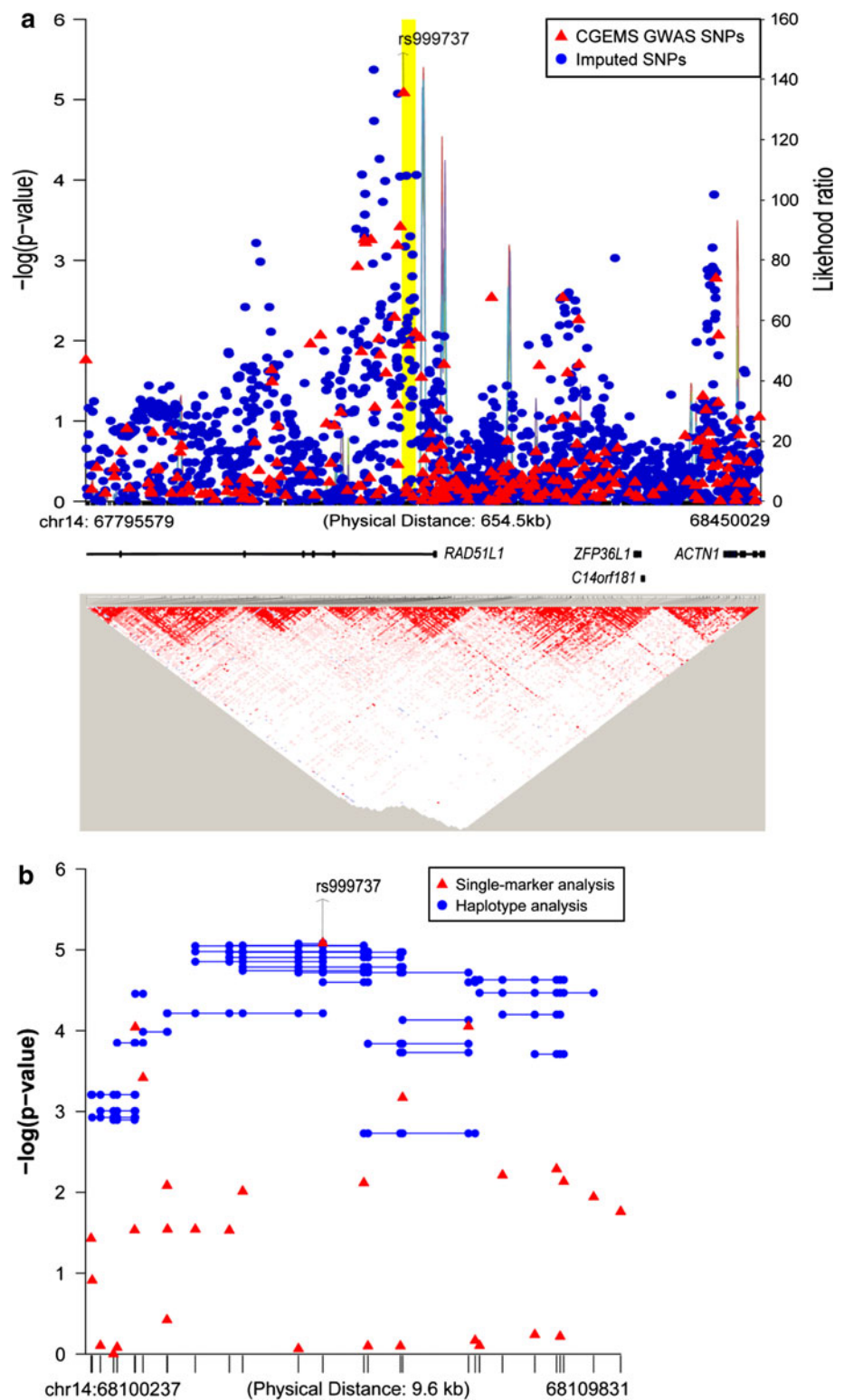
Table 1 Summary of genotyped and imputed SNPs within 3.93 Mb flanking region of 14q24.1

		N (%)
SNP source		
Genotyped	CGEMS phase1 only	733 (6.64)
	CGEMS phase2 only	4 (0.04)
	CGEMS phase 1 and 2	37 (0.33)
Imputed	HapMap 3 only	282 (2.55)
	1000 Genomes only	8,874 (80.33)
	HapMap 3 and 1000 Genomes	1,117 (10.11)
	Total	11,047
IMPUTE2-info score	<0.5	5,816 (56.61)
	0.5–0.75	2,635 (25.65)
	0.76–0.9	1,256 (12.23)
	≥ 0.9	566 (5.51)
	Total	10,273
p value of HWE in controls	$>10^{-3}$	7,726 (75.21)
	10^{-3} – 10^{-4}	494 (4.81)
	10^{-4} – 10^{-5}	334 (3.25)
	10^{-5} – 10^{-6}	270 (2.63)
	$<10^{-6}$	1,449 (14.10)
	Total	10,273
Overall MAF	0–0.01	3,917 (38.13)
	0.01–0.05	2,407 (23.43)
	0.05–0.10	1,448 (14.10)
	0.10–0.20	1,063 (10.35)
	0.20–0.30	646 (6.29)
	0.30–0.40	442 (4.30)
	0.40–0.5	350 (3.41)
	Total	10,273

although 13 of the 45 SNPs generated from imputation that were statistically significant in their association with breast cancer risk were located within the coding region of *RAD51L1*, only 1 SNP showed nominal LD correlation with rs999737 (rs10130694; $r^2 = 0.107$), while others shared minimal pair-wise LD with rs999737 (Table 2). To evaluate whether SNPs across the *RAD51L1* coding region (e.g., in introns and exons) were strongly associated with breast cancer risk, we examined the group of 464 genotyped and imputed SNPs (including rs999737). We discovered that 92 SNPs were located within the *RAD51L1* gene and that 12 showed minimally significant association with breast cancer risk ($P_{\text{adj}} < 0.05$). However, upon conditioning on rs999737, none remained significantly associated with breast cancer risk, thus suggesting that this set of markers tag a single locus.

In addition to the region proximal to rs999737 within *RAD51L1*, we observed several short regions with clusters of SNPs showing nominal association with breast cancer

Fig. 1 Association results, recombination and linkage disequilibrium plots for chromosome 14q21 region. **a** Single-marker analysis of CGEMS GWAS SNPs (red triangles) and imputed SNPs (blue circles) are shown in the top panel with $-\log_{10} P$ values (left y axis). Overlaid on the panel are the likelihood ratio statistics (right y axis) to estimate putative recombination hotspot across the region, based on the 5 sets of 100 randomly selected control samples (connected lines in red, green, blue, purple, and cyan). Pairwise r^2 values based on control samples are displayed at the bottom panel for all SNPs in this analysis. Genomic coordinates are based on NCBI Human Genome Build 36.3. **b** Variable-sized sliding-window haplotype analysis (blue dots and connected lines) and single-marker analysis results of the SNPs (red triangles) are shown with $-\log_{10} P$ values (left y axis) in a 9.6 kb region (yellow highlight in a) that includes SNP rs999737. Genomic coordinates are based on NCBI Human Genome Build 36.3



risk; most importantly, each interval contained in silico genomic elements that could be related to breast cancer (Table 2; see footnotes). The first cluster was 3,527 bp in length (chr14:68,406,986–68,410,513), and contained 6 SNPs that localize to a mammary gland-specific EST,

AK02224. The second cluster of 722 bp (chr14:68,404,595–68,405,317), contained 6 SNPs within a polymorphic insertion/deletion, estimated to be 2,841 bp. The third cluster was the smallest at 396 bp (chr14:68,399,592–68,399,988), and contained only 3 SNPs that mapped to a

Table 2 Forty-six genotyped and imputed SNPs demonstrating nominal significant association with breast cancer risk ($P_{\text{adj}} < 0.05$)

SNP	Position ^a	Source ^b	A1/A2 ^c	MAF CS:CN ^d	OR _{adj} (95%CI) ^e	P_{adj} ^e	r^2 with rs999737
rs17105269	67795579	GWAS-P2	A/T	0.1711:0.1581	1.10 (1.02–1.19)	1.74E–02	0.0069
rs11626138	67856746	1kG	A/G	0.2138:0.2025	1.07 (1.01–1.15)	3.62E–02	0.0208
rs4902567	67870353	1kG	T/C	0.2028:0.1922	1.07 (1.00–1.15)	4.07E–02	0.0186
rs4902587	67975441	1kG	A/G	0.1884:0.1749	1.10 (1.03–1.17)	7.75E–03	0.045
rs2331703	67976519	HM3,1kG	C/T	0.1966:0.1861	1.08 (1.01–1.15)	3.16E–02	0.0312
rs2331704	67976586	GWAS-P1P2	G/A	0.1991:0.1887	1.08 (1.01–1.15)	3.26E–02	0.0295
rs4902588	67978211	1kG	C/T	0.1976:0.1864	1.08 (1.01–1.06)	2.21E–02	0.0302
rs2753403	67982938	HM3,1kG	A/G	0.1955:0.1841	1.08 (1.01–1.06)	2.00E–02	0.0331
rs2525507	68008117	HM3,1kG	G/A	0.2076:0.1959	1.08 (1.01–1.15)	3.38E–02	0.0395
rs2842320	68013826	GWAS-P1	G/A	0.1863:0.1739	1.10 (1.0 2–1.17)	1.10E–02	0.0345
rs999737	68104435	GWAS-P1P2	T/C	0.2127:0.2380	0.87 (0.81–0.92)	8.23E–06	–
rs8009944	68109341	GWAS-P1P2	C/A	0.2834:0.2676	1.08 (1.02–1.15)	1.14E–02	0.0947
rs10130694	68109831	1kG	T/C	0.2787:0.2634	1.08 (1.01–1.15)	1.73E–02	0.1068
rs1290999	68115629	GWAS-P1P2	A/G	0.2133:0.1987	1.09 (1.02–1.17)	8.03E–03	0.063
rs10483818	68189896	GWAS-P1P2	A/G	0.0495:0.0414	1.21 (1.07–1.37)	2.91E–03	0.004
rs2043672	68258829	1kG	G/C	0.0469:0.0559	0.84 (0.74–0.94)	2.88E–03	0
rs2025053	68258941	GWAS-P1P2	A/G	0.0569:0.0665	0.85 (0.74–0.95)	2.91E–03	0.0002
14-68260645	68260645	1kG	T/C	0.0475:0.0563	0.84 (0.75–0.95)	3.89E–03	0.0001
rs8013316	68261417	1kG	C/A	0.0478:0.0564	0.84 (0.75–0.95)	4.52E–03	0.0001
rs55761709	68263324	1kG	T/C	0.0469:0.0560	0.84 (0.74–0.94)	2.79E–03	0.0001
14-68264094	68264094	1kG	G/A	0.047:0.0559	0.84 (0.74–0.94)	2.86E–03	0.0001
rs10483821	68264360	HM3,1kG	G/C	0.0541:0.0625	0.86 (0.77–0.96)	8.89E–03	0
14-68264618	68264618	1kG	G/T	0.0466:0.0558	0.83 (0.74–0.94)	2.59E–03	0.0001
rs8015998	68264784	1kG	C/T	0.0534:0.0621	0.86 (0.77–0.96)	6.22E–03	0
rs1274944	68395187	GWAS-P1P2	G/A	0.1758:0.1655	1.07 (1.00–1.15)	4.90E–02	0.0016
rs440317	68399592	HM3,1kG	T/C	0.3122:0.2943	1.09 (1.02–1.15)	6.15E–03	0.0012
rs1742500	68399654	HM3,1kG	A/G	0.3297:0.3120	1.08 (1.02–1.15)	7.16E–03	0.001
rs453964	68399988	HM3,1kG	T/G	0.3364:0.3188	1.08 (1.02–1.14)	7.01E–03	0.0012
rs445863	68400930	HM3,1kG	T/C	0.3324:0.3122	1.10 (1.04–1.16)	1.57E–03	0.001
rs61987505	68402055	1kG	A/G	0.3159:0.2966	1.09 (1.03–1.16)	2.03E–03	0.0012
rs1742890	68402891	1kG	C/G	0.3314:0.3144	1.08 (1.02–1.15)	1.03E–02	0
rs1742886	68404595	HM3,1kG	G/C	0.3142:0.2941	1.10 (1.04–1.16)	1.20E–03	0.0013
rs1742498	68404615	1kG	G/A	0.3001:0.2779	1.12 (1.05–1.19)	6.95E–04	0.0006
rs10139065	68405176	HM3,1kG	G/A	0.3147:0.2950	1.10 (1.04–1.16)	1.45E–03	0.0012
rs10139069	68405191	1kG	G/A	0.3148:0.2950	1.10 (1.04–1.16)	1.41E–03	0.0012
rs10151103	68405195	1kG	T/C	0.3148:0.2950	1.10 (1.04–1.16)	1.41E–03	0.0012
rs7401911	68405317	HM3,1kG	T/C	0.3147:0.2950	1.10 (1.04–1.16)	1.45E–03	0.0012
rs4902660	68405884	1kG	T/C	0.3333:0.3119	1.10 (1.04–1.17)	1.29E–03	0.0008
rs6573861	68406898	1kG	G/A	0.3148:0.2950	1.10 (1.04–1.16)	1.41E–03	0.0012
rs4902661	68406934	1kG	C/T	0.3271:0.3090	1.09 (1.03–1.15)	5.35E–03	0.0014
rs4902662	68406986	1kG	G/C	0.3182:0.2993	1.09 (1.03–1.16)	2.35E–03	0.0005
rs7149938	68407529	1kG	G/A	0.3303:0.3110	1.09 (1.03–1.16)	2.93E–03	0.0019
rs1475035	68407977	GWAS-P1P2	T/C	0.3145:0.2952	1.10 (1.04–1.16)	1.67E–03	0.0012
rs4312253	68408187	1kG	C/T	0.3149:0.2951	1.10 (1.04–1.16)	1.42E–03	0.0012
rs184156	68410352	1kG	T/C	0.3951:0.4122	0.93 (0.88–0.99)	1.52E–02	0.0033

Table 2 continued

SNP	Position ^a	Source ^b	A1/A2 ^c	MAF CS:CN ^d	OR _{adj} (95%CI) ^e	P_{adj} ^e	r^2 with rs999737
rs1475034	68410513	GWAS-P1	T/G	0.3951:0.4122	1.09 (1.02–1.16)	8.57E–03	0.0005

Pair-wise linkage disequilibrium (r^2 with rs999737) with rs999737 was calculated and no significant LD was shared between rs999737 and any of these SNPs. OR (95%CI) and p-values were estimated from logistic regression with an additive genetic effect, adjusted for age (in 10-year categories), study sites, and 4 significant principle components

^a Genomic coordinates are based on NCBI Human Genome Build 36.3

^b SNP marker sources. GWAS-P1: actual genotyped SNPs in CGEMS GWAS Phase 1; GWAS-P2: actual genotyped SNPs in CGEMS GWAS Phase 2; 1kG: imputed SNPs based on 1000 Genomes Project data; HM3: imputed SNPs based on HapMap Phase 3 data

^c A1 marks the minor allele; A2 marks the major allele

^d Minor allele frequency (MAF) for cases (CS) and controls (CN)

^e Estimates from logistic regression with an additive genetic effect, adjusted for age (in 10-year categories), study sites, and 4 significant principle components

mammary gland-specific EST DA647031. The clusters of SNPs were separated from rs999737 by recombination hotspots, which we estimated based on available controls (Fig. 1a), and the two regions shared minimal LD correlations (Table 2).

Variable-sized sliding window haplotype analysis

Using 5 non-overlapping sets of 100 controls, randomly selected from the entire control set, we calculated and plotted the recombination pattern and LD structure for CGEMS data (Fig. 1a). To further explore whether one or more haplotype can explain the signal, we applied a variable-sized sliding-window haplotype analysis across the imputed region (Table 4; Fig. 1b). Haplotypes inclusive of rs999737 were significantly associated with breast cancer; the strongest association was a 2-SNP haplotype between rs999737 and 14-68103993 (P value = 8.39×10^{-6}). For rs12588940, a SNP that is 3.6 kb away from rs999737, a 7-SNP haplotype inclusive of these two SNPs remained significantly associated with breast cancer (p value = 2.13×10^{-5}). We examined haplotypes within the short clusters of interest, and found that they were moderately associated with breast cancer risk (chr14:68,406,986–68,410,514 interval, smallest P value = 1.36×10^{-3} ; chr14:68,404,595–68,405,317 interval, smallest P value = 5.17×10^{-4} ; chr14:68,399,592–68,399,988 interval, smallest P value = 3.72×10^{-3}). Substantially larger independent follow-up studies will be required to determine if these signals indicate a distinct locus.

Discussion

To follow-up on the original rs999737 signal detected in the CGEMS breast cancer study (Thomas et al. 2009), we have conducted a fine-mapping analysis of the genomic region surrounding rs999737, which localizes to the *RAD51L1* gene in the 14q21 interval (Thomas et al. 2009).

To explore the set of common and uncommon untested variants in this interval, we imputed SNPs based on CGEMS Stages 1 and 2 with data from the HapMap3 and 1000 Genome Project, and applied single-marker and haplotype analyses. Our study has shown that the original SNP, rs999737, retained the strongest association signal for breast cancer susceptibility in this region of 14q21, even after correcting for analysis on the discovery set.

In an analysis of SNPs from across a nearly 4 MB region on 14q24.1, we observed 46 SNPs (inclusive of rs999737) demonstrating a nominal association with breast cancer risk ($P_{adj} < 0.05$); however, the initial discovery SNP, rs999737 remained most significant in Stage 1 and 2 ($P_{adj} = 8.23 \times 10^{-6}$). To evaluate whether any of the signals contributed to breast cancer susceptibility independently from rs999737, we conducted a conditional analysis using the rs999737 discovery marker for the remaining 45 SNPs. While 32 SNPs retained nominal significance for breast cancer risk, in each case, the signal was attenuated. In an exploration of the correlation between each of the 45 SNPs and rs999737 in the control set, we estimated the pair-wise LD (r^2 values). None of the pairs demonstrated a high correlation and we observed residual correlation with one signal marked by rs999737 (Tables 2, 3). In a variable-sized sliding-window haplotype analysis (2–8 SNPs) for all genotyped and imputed SNPs, the results revealed that haplotypes containing rs999737 showed the highest association with breast cancer risk, with P value = 8.39×10^{-6} . Thus, both single-marker and haplotype analysis suggested that the genomic element tagged by rs999737 had the strongest association with breast cancer risk (Thomas et al. 2009).

To determine whether the imputed SNPs exhibited different effects for distinct breast cancer subtypes, specifically, ER and PR subtypes, we conducted a stratified analysis in those cases with reliable ER/PR status (approximately 80%). Supplementary Table 1 shows results for the top 46 SNPs (inclusive of rs999737) with

Table 3 Thirty-two imputed and genotyped SNPs with nominally significant association after sequential conditioning on rs999737 ($P_{\text{adjRS99}} < 0.05$)

SNP	Position ^a	Source ^b	A1/A2 ^c	OR _{adj} (95%CI) ^d	P_{adj}	OR _{adjRS99} (95%CI) ^e	P_{adjRS99}	r^2 with rs999737	Note ^f
rs1742498 ^g	68404615	1kG	G/A	1.12 (1.05–1.19)	6.95E–04	1.11 (1.04–1.18)	1.01E–03	0.0006	rs71824488 (indel); heterozygous deletion for K562 cells
rs4902660	68405884	1kG	T/C	1.10 (1.04–1.17)	1.29E–03	1.10 (1.04–1.17)	1.96E–03	0.0008	Heterozygous deletion for K562 cells
rs1742886 ^g	68404595	HM3, 1kG	G/C	1.10 (1.04–1.16)	1.20E–03	1.09 (1.03–1.16)	2.04E–03	0.0013	rs71824488 (indel)
14-68264618	68264618	1kG	G/T	0.83 (0.74–0.94)	2.59E–03	0.83 (0.74–0.93)	2.05E–03	0.0001	Heterozygous deletion for K562 cells
rs2025053	68258941	GWAS-PIP2	G/A	0.85 (0.74–0.95)	2.91E–03	0.84 (0.76–0.94)	2.14E–03	0.0002	LINE element; heterozygous deletion for K562 cells
rs55761709	68263324	1kG	C/T	0.84 (0.74–0.94)	2.79E–03	0.83 (0.74–0.94)	2.21E–03	0.0001	Heterozygous deletion for K562 cells
14-68264094	68264094	1kG	G/A	0.84 (0.74–0.94)	2.86E–03	0.83 (0.74–0.94)	2.25E–03	0.0001	Heterozygous deletion for K562 cells
rs10139069 ^g	68405191	1kG	G/A	1.10 (1.04–1.16)	1.41E–03	1.09 (1.03–1.15)	2.32E–03	0.0012	rs71824488 (indel); heterozygous deletion for K562 cells
rs10151103 ^g	68405195	1kG	T/C	1.10 (1.04–1.16)	1.41E–03	1.09 (1.03–1.15)	2.32E–03	0.0012	rs71824488 (indel); heterozygous deletion for K562 cells
rs6573861	68406898	1kG	G/A	1.10 (1.04–1.16)	1.41E–03	1.09 (1.03–1.15)	2.32E–03	0.0012	DNA element; heterozygous deletion for K562 cells
rs4312253 ^h	68408187	1kG	C/T	1.10 (1.04–1.16)	1.42E–03	1.09 (1.03–1.15)	2.34E–03	0.0012	Mammary gland EST AK022224
rs2043672	68258829	1kG	C/G	0.84 (0.74–0.94)	2.88E–03	0.83 (0.74–0.94)	2.38E–03	0.0000	LINE element; heterozygous deletion for K562 cells
rs445863	68400930	HM3, 1kG	T/C	1.10 (1.04–1.16)	1.57E–03	1.09 (1.03–1.16)	2.38E–03	0.0010	LTR element; Dnase sensitive; heterozygous deletion for K562 cells
rs10139065 ^g	68405176	HM3, 1kG	G/A	1.10 (1.04–1.16)	1.45E–03	1.09 (1.03–1.16)	2.39E–03	0.0012	rs71824488 (indel); heterozygous deletion for K562 cells
rs7401911 ^g	68405317	HM3, 1kG	T/C	1.10 (1.04–1.16)	1.45E–03	1.09 (1.03–1.16)	2.39E–03	0.0012	rs71824488 (indel); heterozygous deletion for K562 cells
rs1475035 ^h	68407977	GWAS-PIP2	T/C	1.10 (1.04–1.16)	1.67E–03	1.09 (1.03–1.16)	2.79E–03	0.0012	Mammary gland EST AK022224, transcription factor binding sites
14-68260645	68260645	1kG	C/T	0.84 (0.75–0.95)	3.89E–03	0.84 (0.75–0.94)	3.16E–03	0.0001	Transcription binding site; DNase sensitivity; heterozygous deletion for K562 cells
rs4902662 ^h	68406986	1kG	G/C	1.09 (1.03–1.16)	2.35E–03	1.09 (1.03–1.15)	3.17E–03	0.0005	Mammary gland EST AK022224
rs61987505	68402055	1kG	G/A	1.09 (1.03–1.16)	2.03E–03	1.09 (1.03–1.15)	3.30E–03	0.0012	Within LTR element; heterozygous deletion for K562 cells
rs8013316	68261417	1kG	A/C	0.84 (0.75–0.95)	4.52E–03	0.84 (0.75–0.94)	3.63E–03	0.0001	SINE element; heterozygous deletion for K562 cells
rs7149938 ^h	68407529	1kG	G/A	1.09 (1.03–1.16)	2.93E–03	1.09 (1.03–1.15)	5.25E–03	0.0019	SINE element; mammary gland EST AK02224
rs8015998	68264784	1kG	C/T	0.86 (0.77–0.96)	6.22E–03	0.85 (0.76–0.95)	5.36E–03	0.0000	Transcription factor binding site; heterozygous deletion for K562 cells
rs10483818	68189896	GWAS-PIP2	A/G	1.21 (1.07–1.37)	2.91E–03	1.19 (1.05–1.35)	6.61E–03	0.0040	Heterozygous deletion for K562 cells

Table 3 continued

SNP	Position ^a	Source ^b	A1/A2 ^c	OR _{adj} (95%CI) ^d	P _{adj} ^d	OR _{adj} RS99(95%CI) ^e	P _{adj} RS99 ^e	r ² with rs999737	Note ^f
rs10483821	68264360	HM3, IkG	G/C	0.86 (0.77–0.96)	8.89E–03	0.86 (0.77–0.96)	7.77E–03	0.0000	Transcription binding site; DNase sensitivity; heterozygous deletion for K562 cells
rs440317 ⁱ	68399592	HM3, IkG	T/C	1.09 (1.02–1.15)	6.15E–03	1.08 (1.02–1.15)	9.07E–03	0.0012	Mammary gland EST DA647031
rs4902661	68406934	IkG	C/T	1.09 (1.03–1.15)	5.35E–03	1.08 (1.02–1.15)	9.14E–03	0.0014	DNA element; heterozygous deletion for K562 cells
rs1742890	68402891	IkG	C/G	1.08 (1.02–1.15)	1.03E–02	1.08 (1.02–1.15)	9.35E–03	0.0000	SINE element; rs71824488 (indel); heterozygous deletion for K562 cells
rs1742500 ⁱ	68399654	HM3, IkG	A/G	1.08 (1.02–1.15)	7.16E–03	1.08 (1.02–1.14)	9.92E–03	0.0010	SINE element; mammary gland EST DA647031
rs453964 ⁱ	68399988	HM3, IkG	T/G	1.08 (1.02–1.14)	7.01E–03	1.08 (1.02–1.14)	1.01E–02	0.0012	EST from mammary gland DA647031; transcription factor binding sites; heterozygous deletion for K562 cells
rs1475034 ^h	68410513	GWAS-P1	T/G	1.09 (1.02–1.16)	8.57E–03	1.08 (1.02–1.15)	1.15E–02	0.0005	Mammary gland EST AU121625 and AK022224; heterozygous deletion for K562 cells
rs181456 ^h	68410352	IkG	T/C	0.93 (0.88–0.99)	1.52E–02	0.94 (0.89–0.99)	2.89E–02	0.0033	Mammary gland EST AU121625 and AK022224; heterozygous deletion for K562 cells; CpG islands
rs17105269	67795579	GWAS-P2	A/T	1.10 (1.02–1.19)	1.74E–02	1.09 (1.01–1.18)	3.50E–02	0.0069	Heterozygous deletion for K562 cells

OR (95%CI) and p-values were estimated from logistic regression with an additive genetic effect, adjusted for age (in 10-year categories), study sites, 4 significant principle components and rs999737, when applicable. Three clusters of SNPs are found within in silico elements possibly related to breast cancer, and are marked with additional notation

^a Genomic coordinates are based on NCBI Human Genome Build 36.3

^b SNP marker sources. GWAS-P1 actual genotyped SNPs in CGEMS GWAS Phase 1, GWAS-P2 actual genotyped SNPs in CGEMS GWAS Phase 2, *IkG* imputed SNPs based on 1000 Genome Project data, *HM3* imputed SNPs based on HapMap Phase 3 data

^c A1 marks the minor allele; A2 marks the major allele

^d Estimates from logistic regression with an additive genetic effect, adjusted for age (in 10-year categories), study sites, and four significant principle components

^e Estimates from logistic regression with an additive genetic effect, adjusted for age (in 10-year categories), study sites, four significant principle components, and rs999737

^f SNP notation and associated genomic elements

^g Genomic interval containing six SNPs located in a polymorphic insertion/deletion element

^h Genomic interval containing six SNPs located in a mammary gland EST AU121625

ⁱ Genomic interval containing three SNPs located in a mammary gland EST DA647031

Table 4 Top signals from haplotype analysis associated with breast cancer risk

Starting SNP	Starting SNP position	No. of SNPs in sliding window	Haplotype spanned length (bp)	<i>P</i> value ^a
14-68101612	68101612	8	2,823	6.09E–05
rs1468280	68101613	7	2,822	1.22E–04
rs12588940	68101621	7	3,557	2.13E–05
rs2842324	68102125	6	3,053	8.96E–06
rs8016149	68102742	5	2,436	8.83E–06
rs7153476	68102983	7	2,897	1.63E–05
14-68103993	68103993	2	442	8.39E–06
rs999737	68104435	3	817	2.51E–05
rs8007194	68105178	6	2,016	1.86E–03
14-68105252	68105252	4	1,821	1.45E–04
14-68105840	68105840	3	1,233	1.86E–04
rs17756147	68105880	2	1,193	7.38E–05
rs11628293	68107073	2	121	2.52E–05
rs12100794	68107194	7	1,605	2.35E–05
rs6573841	68107274	7	2,067	3.40E–05

^a Estimates from haplotype-specific tests of each haplotype versus all others, adjusted for age (in 10-year categories), study sites, and four significant principle components

nominal association with breast cancer (Table 2) in this study, stratified by each of ER and PR status. Our data suggest that compared with ER-negative tumors, a subset of SNPs had evidence for a stronger association with ER-positive breast cancer than in ER-negative tumors. This might be due to the predominance of ER-positive subjects among breast cancer cases. The analysis stratified by PR status was unremarkable.

As genome-wide association studies can detect disease susceptibility loci throughout multi-stage study design, there might be potential bias introduced by “winner’s curse” effect (Capen et al. 1971) in the SNP selection process of replication studies (Zhong and Prentice 2008). In our analysis, we conducted the fine-mapping in the same set used to discover the tag SNP, rs999737. We recognized that there could be a bias away from the null in favor of the discovered tag SNP (Zhong and Prentice 2008). Here, we calculated the bias-reduced estimators and confidence intervals using a recently proposed method (Zhong and Prentice 2008, 2009) for top SNPs reported in the present study. The bias-adjusted OR and 95% CI were slightly weaker than the unadjusted estimates but maintained good concordance with unadjusted estimates (Supplementary Table 2). In our study of approximately 5,600 cases and 5,600 controls, we had sufficient power to detect variants with minor allele frequency >0.15 and odds ratio greater

than 1.20 assuming a disease prevalence of 12% at the significance level of $P < 5 \times 10^{-6}$ (Supplementary Table 3).

Our results did not identify new strong signals across the large region extending well beyond the recombination hot spots confirmed in our analysis. In this regard, we did not identify a second locus or a substantially better tag SNP than the originally reported SNP, rs999737, suggesting that the common variant signal associated with breast cancer is most likely due to a single locus on chromosome 14q24.1. Functional analysis of the tag SNP and its highly correlated variants will be needed to define the underlying mechanism and to establish the variant(s) that confer the direct association.

In an exploration of a nearly 4 MB interval, we examined the genomic landscape for each of the 32 SNPs noted above in search of possible biological mechanisms that might be worth pursuit, by either laboratory investigation or by extending the replication effort (Table 3). Several clusters of variants map to regions previously related to breast cancer (Fig. 1b; mammary gland specific EST AK022224: chr14:68,406,968–68,410,613; rs71824488: chr14:68,402,332–68,405,672; mammary gland specific EST DA647031: chr14:68,399,325–68,400,050). A cluster of 6 SNPs is within the mammary gland-specific EST, AK02224; a second cluster of 6 SNPs localizes within a polymorphic indel of 2.841 bp on the + strand, marked by the SNP rs71824488; the third cluster of 3 SNPs localizes to another mammary gland-specific EST, DA647031. We also observed possible signals separate from the signal marked by rs999737, but additional studies will be needed to determine if any of these represent independent loci. It is likely that the additional signals either reflect residual LD across this region, still pointing to the single locus marked by rs999737, or that these observations are due to chance.

RAD51L1 is a paralog of the *RAD51* gene, and it forms a complex with *RAD51L2* in the homologous recombination repair pathway through the complex’s ability to catalyze DNA strand-exchange (Hussain et al. 2004; Miller et al. 2002; Sigurdsson et al. 2001). Interestingly, akin to *RAD51L1*, *RAD51L2* was also identified as a breast cancer susceptibility gene (Meindl et al. 2010). In addition to the *RAD51* paralogs, other homologous recombination repair pathways genes have also been associated with breast cancer susceptibility, such as *ATM*, *CHEK2*, *BRCA1* and *BRCA2* (Hollestelle et al. 2010). It is also noteworthy that *RAD51L1* had previously been identified as a fusion gene in myeloid malignancies, thymoma and autoimmune diseases (Nicodeme et al. 2005; Otero et al. 2005). Together, these observations suggest that genetic variation, both germ-line and somatic, within this region could contribute to one or more oncogenic processes.

Acknowledgments The Nurses' Health Studies are supported by NIH grants CA 65725, CA87969, CA49449, CA67262, CA50385 and 5U01CA098233. The authors thank Barbara Egan, Lori Egan, Helena Judge Ellis, Hardeep Ranu, and Pati Soule for assistance, and the participants in the Nurses' Health Studies.

The WHI program is supported by contracts from the National Heart, Lung and Blood Institute, NIH. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at <http://www.whi.org>.

The ACS study is supported by UO1 CA098710. We thank Cari Lichtman for data management and the participants on the CPS-II.

The PLCO study is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. The authors thank Dr Philip Prorok, Division of Cancer Prevention, National Cancer Institute; the Screening Center investigators and staff of the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO). Most importantly, we acknowledge the study participants for their contributions to making this study possible.

References

- Antoniou AC, Kartsonaki C, Sinilnikova OM, Soucy P, McGuffog L, Healey S, Lee A, Peterlongo P, Manoukian S, Peissel B, Zaffaroni D, Cattaneo E, Barile M, Pensotti V, Pasini B, Dolcetti R, Giannini G, Laura Putignano A, Varesco L, Radice P, Mai PL, Greene MH, Andrulis IL, Glendon G, Ozelik H, Thomassen M, Gerdes AM, Kruse TA, Birk Jensen U, Cruger DG, Caligo MA, Laitman Y, Milgrom R, Kaufman B, Paluch-Shimon S, Friedman E, Loman N, Harbst K, Lindblom A, Arver B, Ehrencrona H, Melin B, Nathanson KL, Domchek SM, Rebbeck T, Jakubowska A, Lubinski J, Gronwald J, Huzarski T, Byrski T, Cybulski C, Gorski B, Osorio A, Ramon YCT, Fostira F, Andres R, Benitez J, Hamann U, Hogervorst FB, Rookus MA, Hoening MJ, Nelen MR, van der Luijt RB, van Os TA, van Asperen CJ, Devilee P, Meijers-Heijboer HE, Gomez Garcia EB, Peock S, Cook M, Frost D, Platte R, Leyland J, Gareth Evans D, Lalloo F, Eeles R, Izatt L, Adlard J, Davidson R, Eccles D, Ong KR, Cook J, Douglas F, Paterson J, John Kennedy M, Miedzybrodzka Z, Godwin A, Stoppa-Lyonnet D, Buecher B, Belotti M, Tirapo C, Mazoyer S, Barjhoux L, Lasset C, Leroux D, Faivre L, Bronner M, Prieur F, Nogues C, Rouleau E, Pujol P, Coupier I, Fréney M; CEMO Study Collaborators, Hopper JL, Daly MB, Terry MB, John EM, Buys SS, Yassin Y, Miron A, Goldgar D; Breast Cancer Family Registry, Singer CF, Tea MK, Pfeiler G, Catharina Dressler A, Hansen TV, Jønson L, Ejlersen B, Bjork Barkardottir R, Kirchoff T, Offit K, Piedmonte M, Rodriguez G, Small L, Boggess J, Blank S, Basil J, Azodi M, Ewart Toland A, Montagna M, Tognazzo S, Agata S, Imyanitov E, Janavicius R, Lazaro C, Blanco I, Pharoah PD, Sucheston L, Karlan BY, Walsh CS, Olah E, Bozsik A, Teo SH, Seldon JL, Beattie MS, van Rensburg EJ, Sluiter MD, Diez O, Schmutzler RK, Wappenschmidt B, Engel C, Meindl A, Rühl I, Varon-Mateeva R, Kast K, Deissler H, Niederacher D, Arnold N, Gadzicki D, Schönbuchner I, Caldes T, de la Hoya M, Nevanlinna H, Aittomäki K, Dumont M, Chiquette J, Tischkowitz M, Chen X, Beesley J, Spurdle AB; kConFab investigators, Neuhausen SL, Chun Ding Y, Fredericksen Z, Wang X, Pankratz VS, Couch F, Simard J, Easton DF, Chenevix-Trench G; on behalf of CIMBA (2011) Common alleles at 6q25.1 and 1p11.2 are associated with breast cancer risk for BRCA1 and BRCA2 mutation carriers. *Hum Mol Genet* 20:3304–3321
- Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265
- Capen EC, Clapp RV, Campbell WM (1971) Competitive bidding in high-risk situations. *J Petrol Tech* 23:641–653
- Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB (2010) Rare variants create synthetic genome-wide associations. *PLoS Biol* 8:e1000294
- Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, Hurler ME, McVean GA (2010) A map of human genome variation from population-scale sequencing. *Nature* 467:1061–1073
- Fearnhead P (2006) SequenceLDhot: detecting recombination hotspots. *Bioinformatics* 22:3061–3066
- Figuroa JD, Garcia-Closas M, Humphreys M, Platte R, Hopper JL, Southey MC, Apicella C, Hammett F, Schmidt MK, Broeks A, Tollenaar RA, Van't Veer LJ, Fasching PA, Beckmann MW, Ekici AB, Strick R, Peto J, Dos Santos Silva I, Fletcher O, Johnson N, Sawyer E, Tomlinson I, Kerin M, Burwinkel B, Marme F, Schneeweiss A, Sohn C, Bojesen S, Flyger H, Nordestgaard BG, Benítez J, Milne RL, Ignacio Arias J, Zamora MP, Brenner H, Müller H, Arndt V, Rahman N, Turnbull C, Seal S, Renwick A, Brauch H, Justenhoven C, Brüning T; The GENICA Network, Chang-Claude J, Hein R, Wang-Gohrke S, Dörk T, Schürmann P, Bremer M, Hillemanns P, Nevanlinna H, Heikkinen T, Aittomäki K, Blomqvist C, Bogdanova N, Antonenkova N, Rogov YI, Karstens JH, Bermisheva M, Prokofieva D, Hanafievich Gantcev S, Khusnutdinova E, Lindblom A, Margolin S, Chenevix-Trench G, Beesley J, Chen X; for the kConFab AOCs Management Group, Mannermaa A, Kosma VM, Soini Y, Kataja V, Lambrechts D, Yesilyurt BT, Christiaens MR, Peeters S, Radice P, Peterlongo P, Manoukian S, Barile M, Couch F, Lee AM, Diasio R, Wang X, Giles GG, Severi G, Baglietto L, Maclean C, Offit K, Robson M, Joseph V, Gaudet M, John EM, Winqvist R, Pyrkäs K, Jukkola-Vuorinen A, Grip M, Andrulis I, Knight JA, Marie Mulligan A, O'Malley FP, Brinton LA, Sherman ME, Lissowska J, Chanock SJ, Hooning M, Martens JW, van den Ouweland AM, Collée JM, Hall P, Czene K, Cox A, Brock IW, Reed MW, Cross SS, Pharoah P, Dunning AM, Kang D, Yoo KY, Noh DY, Ahn SH, Jakubowska A, Lubinski J, Jaworska K, Durda K, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Shen CY, Ding SL, Hsu HM, Yu JC, Anton-Culver H, Ziogas A, Ashworth A, Swerdlow A, Jones M, Orr N, Trentham-Dietz A, Egan K, Newcomb P, Titus-Ernstoff L, Easton D, Spurdle AB (2011) Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. *Hum Mol Genet* [Epub ahead of print]
- Ghousaini M, Pharoah PD (2009) Polygenic susceptibility to breast cancer: current state-of-the-art. *Future Oncol* 5:689–701
- Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250:1684–1689
- Havre PA, Rice MC, Noe M, Kmiec EB (1998) The human REC2/RAD51B gene acts as a DNA damage sensor by inducing G1 delay and hypersensitivity to ultraviolet irradiation. *Cancer Res* 58:4733–4739
- Hollestelle A, Wasielewski M, Martens JW, Schutte M (2010) Discovering moderate-risk breast cancer susceptibility genes. *Curr Opin Genet Dev* 20:268–276
- Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5:e1000529
- Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K,

- Chatterjee N, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover RN, Thomas G, Chanock SJ (2007) A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 39:870–874
- Hussain S, Wilson JB, Medhurst AL, Hejna J, Witt E, Ananth S, Davies A, Masson JY, Moses R, West SC, de Winter JP, Ashworth A, Jones NJ, Mathew CG (2004) Direct interaction of *FANCD2* with *BRCA2* in DNA damage response pathways. *Hum Mol Genet* 13:1241–1248
- Jemal A, Siegel R, Xu J, Ward E (2010) Cancer statistics, 2010. *CA Cancer J Clin* 60:277–300
- Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skythe A, Hemminki K (2000) 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
- Luna A, Nicodemus KK (2007) *snp.plotter*: an R-based SNP/haplotype association and linkage disequilibrium plotting package. *Bioinformatics* 23:774–776
- Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, Freund M, Lichtner P, Hartmann L, Schaal H, Ramser J, Honisch E, Kubisch C, Wichmann HE, Kast K, Deissler H, Engel C, Muller-Myhsok B, Neveling K, Kiechle M, Mathew CG, Schindler D, Schmutzler RK, Hanenberg H (2010) Germline mutations in breast and ovarian cancer pedigrees establish *RAD51C* as a human cancer susceptibility gene. *Nat Genet* 42:410–414
- Miller KA, Yoshikawa DM, McConnell IR, Clark R, Schild D, Albala JS (2002) *RAD51C* interacts with *RAD51B* and is central to a larger protein complex in vivo exclusive of *RAD51*. *J Biol Chem* 277:8406–8411
- Nicodeme F, Geffroy S, Conti M, Delobel B, Soenen V, Gardel N, Porte H, Copin MC, Lai JL, Andrieux J (2005) Familial occurrence of thymoma and autoimmune diseases with the constitutional translocation $t(14;20)(q24.1;p12.3)$. *Genes Chromosomes Cancer* 44:154–160
- Odero MD, Grand FH, Iqbal S, Ross F, Roman JP, Vizmanos JL, Andrieux J, Lai JL, Calasanz MJ, Cross NC (2005) Disruption and aberrant expression of *HMG2* as a consequence of diverse chromosomal translocations in myeloid malignancies. *Leukemia* 19:245–252
- Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* 42:570–575
- Pruthi S, Gostout BS, Lindor NM (2010) Identification and management of women with *BRCA* mutations or hereditary predisposition for breast and ovarian cancer. *Mayo Clin Proc* 85:1111–1120
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Sigurðsson S, Van Komen S, Bussen W, Schild D, Albala JS, Sung P (2001) Mediator function of the human *Rad51B*-*Rad51C* complex in *Rad51*/RPA-catalyzed DNA strand exchange. *Genes Dev* 15:3308–3318
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989
- Takaku M, Machida S, Hosoya N, Nakayama S, Takizawa Y, Sakane I, Shibata T, Miyagawa K, Kurumizaka H (2009) Recombination activator function of the novel *RAD51*- and *RAD51B*-binding protein, human *EVL*. *J Biol Chem* 284:14326–14336
- Takata M, Sasaki MS, Sonoda E, Fukushima T, Morrison C, Albala JS, Swagemakers SM, Kanaar R, Thompson LH, Takeda S (2000) The *Rad51* paralog *Rad51B* promotes homologous recombinational repair. *Mol Cell Biol* 20:6476–6482
- Thacker J, Zdzienicka MZ (2004) The *XRCC* genes: expanding roles in DNA double-strand break repair. *DNA Repair (Amst)* 3:1081–1090
- The International HapMap3 Consortium (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467:52–58
- Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, Gonzalez-Bosquet J, Prokunina-Olsson L, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver R, Prentice R, Jackson R, Kooperberg C, Chlebowski R, Lissowska J, Peplonska B, Brinton LA, Sigurdson A, Doody M, Bhatti P, Alexander BH, Buring J, Lee IM, Vatten LJ, Hveem K, Kumle M, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover RN, Chanock SJ, Hunter DJ (2009) A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (*RAD51L1*). *Nat Genet* 41:579–584
- Varghese JS, Easton DF (2010) Genome-wide association studies in common cancers—what have we learnt? *Curr Opin Genet Dev* 20:201–209
- Zeidler M, Varambally S, Cao Q, Chinnaiyan AM, Ferguson DO, Merajver SD, Kleer CG (2005) The Polycomb group protein *EZH2* impairs DNA repair in breast epithelial cells. *Neoplasia* 7:1011–1019
- Zhong H, Prentice RL (2008) Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics* 9:621–634
- Zhong H, Prentice RL (2009) Correcting “winner’s curse” in odds ratios from genomewide association findings for major complex human diseases. *Genet Epidemiol* 34:78–91