Published in final edited form as:
Cancer Epidemiol Biomarkers Prev. 2010 April ; 19(4): 1140–1143. doi:10.1158/1055-9965.EPI-10-0085.

A genome-wide association study of prognosis in breast cancer

Elizabeth M. Azzato[1,4,*], **Paul D.P. Pharoah**[1], **Patricia Harrington**[1], **Douglas F. Easton**[2], **David Greenberg**[3], **Neil E. Caporaso**[4], **Stephen J. Chanock**[4], **Robert N. Hoover**[4], **Gilles Thomas**[4], **David J. Hunter**[5], and **Peter Kraft**[5]

[1]Department of Oncology, University of Cambridge, Worts Causeway, Cambridge CB1 8RN, United Kingdom

[2]Department of Oncology, Public Health and Primary Care, Strangeways Research Laboratory, University of Cambridge, Worts Causeway, Cambridge CB1 8RN, United Kingdom

[3]Eastern Cancer Registration and Information Centre, Unit C-Magog Court, Shelford Bottom Cambridge CB22 3AD, United Kingdom

[4]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, DHHS, Bethesda, MD, USA

[5]Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

Abstract

Background—Traditional clinicopathological features of breast cancer do not account for all the variation in survival. Germline genetic variation may provide additional prognostic information.

Materials and Methods—We conducted a GWAS study of survival after a diagnosis of breast cancer by obtaining follow-up data and genotyping information on 528,252 SNPs for 1,145 postmenopausal women with invasive breast cancer (7,711 person years at risk) from the Nurses' Health Study scanned in the Cancer Genetic Markers of Susceptibility initiative. We genotyped the ten most statistically significant loci (most significant SNP located in *ARHGAP10*, $p = 2.28 \times 10^{-7}$) in 4,335 women diagnosed with invasive breast cancer (38,148 years at risk) in the SEARCH breast cancer study.

Results—None of the loci replicated in the SEARCH study (all $p > 0.10$). Assuming a minimum of ten associated loci, the power to detect at least one with a minor allele frequency of 0.2 conferring a relative hazard of 2.0 at genome-wide significance (5×10^{-8}) was 99 percent.

Conclusions—We did not identify any common germline variants associated with breast cancer survival overall.

Impact—Our data suggest it is unlikely that there are common germline variants with large effect sizes for breast cancer survival overall (HR>2). Instead, it is plausible that common variants associated with survival could be specific to tumor subtypes or treatment approaches. New studies, sufficiently powered, are needed to discover new regions associated with survival overall or by subtype or treatment subgroups.

Keywords

breast cancer; prognosis; genome-wide association study; molecular diagnosis and prognosis; single nucleotide polymorphism

[*]Corresponding author: Elizabeth Azzato, M.P.H., 6120 Executive Plaza, Rm 7106, Rockville, MD 20852, USA, Telephone: 301-435-7614, Fax: 301-402-4489, azzatoe2@mail.nih.gov

INTRODUCTION

Traditional clinical and pathological features related to breast cancer prognosis are not adequate predictors of survival (1) and it is possible that variation in germline DNA may provide additional information. Previous studies have focused on candidate genes, relying on our incomplete knowledge of tumor and host biology (2-8). In this study, we aimed to identify common germline variants associated with breast cancer specific survival after a diagnosis of breast cancer using a genome-wide approach. We obtained follow-up information on 1,145 women with invasive post-menopausal breast cancer from the Nurses' Health Study cohort genotyped using the Illumina HumanHap500 array, as part of the Cancer Genetic Markers of Susceptibility (CGEMS) initiative. We genotyped the most statistically significant associations in the Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH) breast cancer study.

METHODS

Sample

The 1,145 women of the NHS/CGEMS sample used in this analysis and the genome wide association study (GWAS) genotyping methods have been previously described (9,10). Briefly, the NHS is a longitudinal study of 121,700 women enrolled in 1976. The CGEMS case-control study is derived from 32,826 participants who provided a blood sample in 1989-1990 and followed for incident breast cancer until May 2004. Follow-up was conducted by personal mailings and searches of the National Death Index. The 1,145 women were genotyped using the the Illumina HumanHap500 array as the first stage of a 3 stage GWAS of breast cancer susceptibility. We removed SNPs with a minor allele frequency (MAF) <1% for a total of 528,252 SNPs.

The SEARCH breast cancer study ascertainment and follow-up is described elsewhere (4). Briefly, SEARCH is an ongoing population-based study of women diagnosed with breast cancer in the region of England included in the Eastern Cancer Registration and Information Centre (ECRIC). Study eligibility includes those diagnosed with invasive breast cancer <70 years of age at the start of the study in 1996, or those diagnosed at age 55 or younger since 1991 and alive at the start of the study (prevalent cases). Follow-up was obtained by death certificate flagging through the Office of National Statistics and the National Health Service Strategic Tracing Service. Genotyping sets 1 and 2 (4,335 women) were included in this analysis. Genotyping was performed using optimized Taqman assays (Applied Biosystems) (11).

All participants in the studies provided informed consent. The NHS/CGMEMs study protocol was approved by the Brigham and Women's Hospital IRB. The SEARCH study was approved by the Eastern Multicentre Research Ethics Committee.

Statistical methods

We fitted Cox proportional hazards models to assess association of genotype with breast cancer specific mortality, adjusting for age category at diagnosis (44-59, 60-69, 70-83). Follow-up for NHS ended at date of death from breast cancer or June 30, 2004. Statistical significance was based on a one degree of freedom trend test.

The ten most statistically significant loci were genotyped in SEARCH and assessed for association with breast cancer specific mortality adjusting for age category at diagnosis (<44, 44-59, 60-69), using Cox proportional hazards models allowing for left truncated (prevalent case) data (12). Follow-up ended for SEARCH at date of death from breast cancer or November

30, 2006; follow-up was censored at ten years. Since the SEARCH population includes some pre-menopausal cases, analyses limited to individuals with an age at diagnosis ≥44 and ≥55 years were also performed.

RESULTS

The 1,145 women participating in the NHS/CGEMS study provided 7,711 person-years at risk (93 breast cancer deaths) (Table 1). We tested the 528,252 SNPs for an association with breast cancer survival and no association reached nominal GWAS significance threshold (5×10^{-8}). The most significant locus was in the *ARHGAP10* gene (2.3×10^{-7}).

The ten loci were genotyped in up to 4,335 women participating in the SEARCH study, who provided 28,148 years at risk (587 breast cancer deaths). None of the loci replicated in the SEARCH study (Table 2). There were no significant differences in the analyses when stratified by age at diagnosis ≥44 and ≥55 years (data not shown).

DISCUSSION

The NHS/CGEMS study provides the unique opportunity for an agnostic search of the genome for common genetic variants associated with breast cancer prognosis. To date, this is the first GWAS of breast cancer survival. However, we did not observe any SNP associations with a genome level of statistical significance (5×10^{-8}), nor did we replicate any of the ten most statistically significant loci discovered in the GWAS in the SEARCH study.

Assuming a minimum of ten associated loci, power to detect at least one where the risk allele frequency is 0.2 conferring relative hazards of 1.6, 1.8 and 2.0 at genome-wide significance (5×10^{-8}), taking into account the staged study design, was 48, 89 and 99 percent, respectively. As power to detect larger magnitude effects and more prevalent alleles is correspondingly greater, it is unlikely that we have missed common variants with large effect sizes. However, it is possible that variants with more modest effects were missed in the discovery analysis and were not carried forward to the replication phase. Also, our power in the discovery GWAS is less favorable for rare variants or genes acting via a recessive mechanism.

Breast cancer is a heterogeneous disease and its prognosis varies significantly across tumor subtypes (13-18); additional factors such as patient characteristics (e.g., age, co-morbidity, diet, etc), treatment regimen, compliance and individual pharmacogenetics also impact survival (19). It is plausible germline genetic variation could be associated with survival by tumor subtype or treatment approaches. However, the power to investigate subgroups as well as interactions with environmental factors is limited with the current data set and will require larger consortial studies. Furthermore, the determination of common variants associated with survival is challenging in studies designed to discover common variants for etiology because of issues related to study design.

In conclusion, our study suggests that it is unlikely that there are many common germline variants with large effects (HR>2) on general breast cancer survival. Further candidate gene and GWA studies powered for common variants with modest effects on survival, as well as tumor and treatment subgroups, are required.

Acknowledgments

We thank Craig Luccarini, Don Conroy and the SEARCH team. The Nurse's Health Study is supported by US NIH grant P01 CA087969. SEARCH is funded through a program grant from Cancer Research UK. EMA is funded through the NIH-University of Cambridge Graduate Partnership Program and the Intramural research program of the National Cancer Institute.

References

1. Tavassoli, FA.; Devilee, P. Pathology and Genetics of Tumours of the Breast and Female Genital Organs. International Agency for Research on Cancer, Oxford University Press; Lyon: 2003.
2. Udler M, Azzato EM, Healey CS, Ahmed S, Pooley K, Greenberg D, Shah M, Teschendorff AE, Caldas C, Dunning A, Ostrander EA, Caporaso N, Easton D, Pharoah P. Common germline polymorphisms in COMT, CYP19A1, ESR1, PGR, SULT1E1 and STS and survival after a diagnosis of breast cancer. IJC 2009;125:2687–96.
3. Abraham JE, Harrington P, Driver KE, Tyrer J, Easton DF, Dunning AM, Pharoah PD. Common polymorphisms in the prostaglandin pathway genes and their association with breast cancer susceptibility and survival. Clin Cancer Res 2009;15:2181–91. [PubMed: 19276290]
4. Azzato EM, Driver KE, Lesueur F, Shah M, Greenberg D, Easton DF, Teschendorff AE, Caldas C, Caporaso NE, Pharoah PD. Effects of common germline genetic variation in cell cycle control genes on breast cancer survival: results from a population-based cohort. Breast Cancer Res 2008;10:R47. [PubMed: 18507837]
5. Fagerholm R, Hofstetter B, Tommiska J, Aaltonen K, Vrtel R, Syrjakoski K, Kallioniemi A, Kilpivaara O, Mannermaa A, Kosma VM, Uusitupa M, Eskelinen M, Kataja V, Aittomaki K, von Smitten K, Heikkila P, Lukas J, Holli K, Bartkova J, Blomqvist C, Bartek J, Nevanlinna H. NAD(P)H:quinone oxidoreductase 1 NQO1*2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer. Nat Genet 2008;40:844–53. [PubMed: 18511948]
6. Fasching PA, Loehberg CR, Strissel PL, Lux MP, Bani MR, Schrauder M, Geiler S, Ringleff K, Oeser S, Weihbrecht S, Schulz-Wendtland R, Hartmann A, Beckmann MW, Strick R. Single nucleotide polymorphisms of the aromatase gene (CYP19A1), HER2/neu status, and prognosis in breast cancer patients. Breast Cancer Res Treat 2008;112:89–98. [PubMed: 18049890]
7. Udler M, Maia AT, Cebrian A, Brown C, Greenberg D, Shah M, Caldas C, Dunning A, Easton D, Ponder B, Pharoah P. Common germline genetic variation in antioxidant defense genes and survival after diagnosis of breast cancer. J Clin Oncol 2007;25:3015–23. [PubMed: 17634480]
8. Goode EL, Dunning AM, Kuschel B, Healey CS, Day NE, Ponder BA, Easton DF, Pharoah PP. Effect of germ-line genetic variation on breast cancer survival in a population-based study. Cancer Res 2002;62:3052–7. [PubMed: 12036913]
9. 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. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–4. [PubMed: 17529973]
10. 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. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet 2009;41:579–84. [PubMed: 19330030]
11. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo KY, Noh DY, Ahn SH, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low YL, Bogdanova N, Schurmann P, Dork T, Tollenaar RA, Jacobi CE, Devilee P, Klijn JG, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, MacPherson G, Reed MW, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justenhoven C, Brauch H, Hamann U, Ko YD, Spurdle AB, Beesley J, Chen X, Mannermaa A,

Kosma VM, Kataja V, Hartikainen J, Day NE, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–93. [PubMed: 17529967]

12. Azzato EM, Greenberg D, Shah M, Blows F, Driver KE, Caporaso NE, Pharoah PD. Prevalent cases in observational studies of cancer survival: do they bias hazard ratio estimates? Br J Cancer 2009;100:1806–1811. [PubMed: 19401693]
13. Anderson WF, Jatoi I, Devesa SS. Distinct breast cancer incidence and prognostic patterns in the NCI's SEER program: suggesting a possible link between etiology and outcome. Breast Cancer Res Treat 2005;90:127–37. [PubMed: 15803359]
14. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999–2009. [PubMed: 12490681]
15. Naderi A, Teschendorff AE, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG, Robertson JF, Aparicio S, Ellis IO, Brenton JD, Caldas C. A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene 2007;26:1507–1516. [PubMed: 16936776]
16. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–6. [PubMed: 11823860]
17. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P, Borresen-Dale AL. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74. [PubMed: 11553815]
18. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100:8418–23. [PubMed: 12829800]
19. Marsh S, McLeod HL. Pharmacogenetics and oncology treatment for breast cancer. Expert Opin Pharmacother 2007;8:119–27. [PubMed: 17257083]

Table 1

Study population characteristics

	NHS/CGEMS		SEARCH	
Total Number of Subjects	1,145		4,335	
Total time at risk (years)	7,711		28,148*	
Median time at risk (years)	6.00	(0.001 - 15.00)†	6.51	(0.036 - 9.81)†
Number of breast cancer deaths	93		587	
Annual mortality rate	0.012	(0.010 - 0.014)‡	0.021	(0.019 - 0.023) ‡
5-year survival rate	0.94	(0.944-0.969) ‡	0.89	(0.883 - 0.905) ‡
Median age at diagnosis, years	66	(44-83) †	51	(23-69) †
Histopathological grade				
Well differentiated	209	18.25%	858	19.79%
Moderately differentiated	389	33.97%	1,652	38.11%
Poorly differentiated	243	21.22%	989	22.81%
Unknown	304	26.55%	836	19.28%
Clinical Stage				
I	725	63.32%	2,133	49.20%
II	300	26.20%	1,924	44.38%
III	49	4.28%	142	3.28%
IV	0	0.00%	44	1.01%
Unknown	71	6.20%	92	2.12%
ER status				
Positive	807	70.48%	2,440	56.29%
Negative	181	15.81%	639	14.74%
Unknown	157	13.71%	1,256	28.97%

*Follow-up censored at 10 years, analysis allowing for left-truncated data

†Range of variable

‡95% confidence interval

Table 2

Hazard ratios, 95% confidence intervals and p-values for NHS/CGEMS and SEARCH studies for the top ten most significant loci from NHS/CGEMS breast cancer survival GWAS

			NHS/CGEMS			SEARCH		
SNP	**Location* Nearby Gene(s)**	**Alleles†**	**MAF**	**HR‡ (95% CI)**	**p**	**MAF**	**HR§ (95% CI)**	**p**
rs13124167	4q31.23 148894643 *ARHGAP10*	T,C	0.12	2.48 (2.13 – 2.82)	2.28×10^{-7}	0.11	1.00 (0.83 – 1.19)	0.97
rs4529739	1p32.1 60477371	T,C	0.09	2.31 (1.96 – 2.65)	2.69×10^{-6}	0.11	0.99 (0.83 – 1.19)	0.95
rs11591508	10p11.22 33324659	C,T	0.06	2.85 (2.41 – 3.29)	3.27×10^{-6}	0.06	0.91 (0.71 – 1.16)	0.43
rs2571236	18q21.31 53607672	G,A	0.23	1.96 (1.67 – 2.25)	5.32×10^{-6}	0.23	0.99 (0.86 – 1.13)	0.84
rs3094663	6p21.33 31215066 *PSORS1C1, CDSN, PSORS1C2, C6orf18*	G,A	0.30	1.94 (1.64 – 2.24)	1.10×10^{-5}	0.28	0.98 (0.81 – 1.18)	0.82
rs352457	15q22.31 63564258 *DPP8*	G,A	0.06	2.46 (2.05 – 2.87)	1.82×10^{-5}	0.06	1.06 (0.83 – 1.35)	0.64
rs936503	18q23 74788302	T,C	0.39	0.49 (0.15 – 0.82)	2.49×10^{-5}	0.37	1.08 (0.95 – 1.21)	0.23
rs2282079	9p13.2 37026247 *PAX5, LOC401504*	G,A	0.05	2.03 (1.70 – 2.37)	2.76×10^{-5}	0.04	0.84 (0.61 – 1.16)	0.28
rs17299684	15q25.2 82495353 *ADAMTSL3*	A,G	0.15	1.95 (1.64 – 2.27)	3.25×10^{-5}	0.15	0.87 (0.74 – 1.03)	0.11
rs17296289	10p11.22 33300705 *ITGB1*	G,A	0.06	2.51 (2.06 – 2.96)	5.40×10^{-5}	0.07	0.93 (0.73 – 1.18)	0.53

MAF = minor allele frequency

* dbSNP build 130

† major, minor allele

‡ Adjusted for age at diagnosis categories (44-59, 60-69, 70+)

§ Adjusted for age at diagnosis categories (<44, 44-59, 60-69)