A Genome-Wide Association Study of Prognosis in Breast Cancer

Elizabeth M. Azzato1,4, Paul D.P. Pharoah1, Patricia Harrington1, Douglas F. Easton2, David Greenberg3, Neil E. Caporaso4, Stephen J. Chanock4, Robert N. Hoover4, Gilles Thomas4, David J. Hunter5, and Peter Kraft5

Abstract

Background: Traditional clinicopathologic 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 genome-wide association study of survival after a diagnosis of breast cancer by obtaining follow-up data and genotyping information on 528,252 single-nucleotide polymorphisms 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 10 most statistically significant loci (most significant single-nucleotide polymorphism 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 (Studies of Epidemiology and Risk factors in Cancer Heredity) breast cancer study.

Results: None of the loci replicated in the SEARCH study (all \( P > 0.10 \)). Assuming a minimum of 10 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 (\( P = 5 \times 10^{-8} \)) was 99%.

Conclusion: We did not identify any common germline variants associated with breast cancer survival overall.

Impact: Our data suggest that it is unlikely that there are common germline variants with large effect sizes for breast cancer survival overall (hazard ratio >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. Cancer Epidemiol Biomarkers Prev; 19(4); 1140–3. ©2010 AACR.

Introduction

Traditional clinical and pathologic 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 postmenopausal breast cancer from the Nurses’ Health Study (NHS) 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 SEARCH (Studies of Epidemiology and Risk factors in Cancer Heredity) breast cancer study.

Materials and 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 were 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 Illumina HumanHap500 array as the first stage of a three-stage GWAS of breast cancer susceptibility. We removed single-nucleotide polymorphisms (SNP) with a minor allele frequency of <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. Study eligibility includes those diagnosed with invasive breast cancer at age <70 y 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 done using optimized TaqMan assays (Applied Biosystems; ref. 11).

All participants in the studies provided informed consent. The NHS/CGEMS study protocol was approved by the Brigham and Women’s Hospital Institutional Review Board. The SEARCH study was approved by the Eastern Multicentre Research Ethics Committee.

### Statistical methods

We fitted Cox proportional hazards models to assess the association of genotype with breast cancer–specific mortality, adjusting for age category at diagnosis (44-59, 60-69, and 70-83 y). Follow-up for NHS ended at date of death from breast cancer or June 30, 2004. Statistical significance was based on a 1-degree-of-freedom trend test.

The 10 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, and 60-69 y), 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 10 y. Because the SEARCH population includes some premenopausal cases, analyses limited to individuals with ages at diagnosis of ≥44 and ≥55 y 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

### Table 1. Study population characteristics

<table>
<thead>
<tr>
<th></th>
<th>NHS/CGEMS</th>
<th>SEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total no. of subjects</td>
<td>1,145</td>
<td>4,335</td>
</tr>
<tr>
<td>Total time at risk, y</td>
<td>7,711</td>
<td>28,148*</td>
</tr>
<tr>
<td>Median time at risk, y</td>
<td>6.00 (0.001-15.00)†</td>
<td>6.51 (0.036-9.81)†</td>
</tr>
<tr>
<td>No. of breast cancer deaths</td>
<td>93</td>
<td>587</td>
</tr>
<tr>
<td>Annual mortality rate</td>
<td>0.012 (0.010-0.014)‡</td>
<td>0.021 (0.019-0.023)‡</td>
</tr>
<tr>
<td>5-y survival rate</td>
<td>0.94 (0.944-0.969)‡</td>
<td>0.89 (0.883-0.905)‡</td>
</tr>
<tr>
<td>Median age at diagnosis, y</td>
<td>66 (44-83)†</td>
<td>51 (23-69)†</td>
</tr>
<tr>
<td>Histopathologic grade, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well differentiated</td>
<td>209 (18.25)</td>
<td>858 (19.79)</td>
</tr>
<tr>
<td>Moderately differentiated</td>
<td>389 (33.97)</td>
<td>1,652 (38.11)</td>
</tr>
<tr>
<td>Poorly differentiated</td>
<td>243 (21.22)</td>
<td>989 (22.81)</td>
</tr>
<tr>
<td>Unknown</td>
<td>304 (26.55)</td>
<td>836 (19.28)</td>
</tr>
<tr>
<td>Clinical stage, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>725 (63.32)</td>
<td>2,133 (49.20)</td>
</tr>
<tr>
<td>II</td>
<td>300 (26.20)</td>
<td>1,924 (44.38)</td>
</tr>
<tr>
<td>III</td>
<td>49 (4.28)</td>
<td>142 (3.28)</td>
</tr>
<tr>
<td>IV</td>
<td>0 (0.00)</td>
<td>44 (1.01)</td>
</tr>
<tr>
<td>Unknown</td>
<td>71 (6.20)</td>
<td>92 (2.12)</td>
</tr>
<tr>
<td>Estrogen receptor status, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>807 (70.48)</td>
<td>2,440 (56.29)</td>
</tr>
<tr>
<td>Negative</td>
<td>181 (15.81)</td>
<td>639 (14.74)</td>
</tr>
<tr>
<td>Unknown</td>
<td>157 (13.71)</td>
<td>1,256 (28.97)</td>
</tr>
</tbody>
</table>

*Follow-up censored at 10 y, analysis allowing for left-truncated data.
†Range of variable.
‡95% confidence interval.
an association with breast cancer survival, and no association reached nominal GWAS significance threshold ($P = 5 \times 10^{-8}$). The most significant locus was in the ARHGAP10 gene ($P = 2.3 \times 10^{-7}$).

The 10 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 ages at diagnosis of ≥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 ($P = 5 \times 10^{-8}$), nor did we replicate any of the 10 most statistically significant loci discovered in the GWAS in the SEARCH study.

Assuming a minimum of 10 associated loci, the 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 ($5 \times 10^{-8}$), taking into account the staged study design, was 48%, 89%, and 99%, respectively. Because the power to detect larger magnitude effects and more prevalent alleles is correspondingly greater, it is unlikely that we have missed

<table>
<thead>
<tr>
<th>SNP</th>
<th>Location*</th>
<th>Alleles†</th>
<th>NHS/CGEMS HR‡ (95% CI)</th>
<th>NHL/CGEMS P</th>
<th>SEARCH HR§ (95% CI)</th>
<th>SEARCH P</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs13124167</td>
<td>4q31.23</td>
<td>T, C</td>
<td>2.48 (2.13-2.82)</td>
<td>2.28 × 10⁻⁷</td>
<td>1.00 (0.83-1.19)</td>
<td>0.97</td>
</tr>
<tr>
<td>rs4529739</td>
<td>1p32.1</td>
<td>T, C</td>
<td>2.31 (1.96-2.65)</td>
<td>2.69 × 10⁻⁶</td>
<td>0.99 (0.83-1.19)</td>
<td>0.95</td>
</tr>
<tr>
<td>rs11591508</td>
<td>10p11.22</td>
<td>C, T</td>
<td>2.85 (2.41-3.29)</td>
<td>3.27 × 10⁻⁶</td>
<td>0.91 (0.71-1.16)</td>
<td>0.43</td>
</tr>
<tr>
<td>rs2571236</td>
<td>10q21.31</td>
<td>A, T</td>
<td>1.96 (1.67-2.25)</td>
<td>5.32 × 10⁻⁶</td>
<td>0.99 (0.86-1.13)</td>
<td>0.84</td>
</tr>
<tr>
<td>rs3094663</td>
<td>6p21.33</td>
<td>A, T</td>
<td>1.94 (1.64-2.24)</td>
<td>1.10 × 10⁻⁵</td>
<td>0.98 (0.81-1.18)</td>
<td>0.82</td>
</tr>
<tr>
<td>rs352457</td>
<td>15q22.31</td>
<td>G, A</td>
<td>2.46 (2.05-2.87)</td>
<td>1.82 × 10⁻⁵</td>
<td>1.06 (0.83-1.35)</td>
<td>0.64</td>
</tr>
<tr>
<td>rs93503</td>
<td>15q21.32</td>
<td>G, A</td>
<td>1.95 (1.64-2.27)</td>
<td>3.25 × 10⁻⁵</td>
<td>0.87 (0.74-1.03)</td>
<td>0.11</td>
</tr>
<tr>
<td>rs2282079</td>
<td>30q26.247</td>
<td>A, G</td>
<td>2.00 (1.70-2.37)</td>
<td>2.76 × 10⁻⁵</td>
<td>0.84 (0.61-1.16)</td>
<td>0.28</td>
</tr>
<tr>
<td>rs1729684</td>
<td>15q22.2</td>
<td>A, G</td>
<td>1.95 (1.64-2.27)</td>
<td>3.25 × 10⁻⁵</td>
<td>0.87 (0.74-1.03)</td>
<td>0.11</td>
</tr>
<tr>
<td>rs17296289</td>
<td>10p11.22</td>
<td>C, T</td>
<td>2.51 (2.06-2.96)</td>
<td>5.40 × 10⁻⁵</td>
<td>0.93 (0.73-1.18)</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Abbreviations: MAF, minor allele frequency; HR, hazard ratio; 95% CI, 95% confidence interval.

*dbSNP build 130.
†Major allele, minor allele.
‡Adjusted for age at diagnosis categories (44-59, 60-69, and 70+ y).
§Adjusted for age at diagnosis categories (<44, 44-59, and 60-69 y).
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. In addition, 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, comorbidity, diet, etc.), treatment regimen, compliance, and individual pharmacogenetics also affect survival (19). It is plausible that 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 (hazard ratio >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.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Craig Luccarini, Don Conroy, and the SEARCH team.

Grant Support

The Nurses’ Health Study is supported by U.S. NIH grant P01 CA087969. SEARCH is funded through a program grant from Cancer Research UK. E.M. Azzato is funded through the NIH-University of Cambridge Graduate Partnership Program and the Intramural Research Program of the National 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.

Received 01/19/2010; accepted 01/25/2010; published OnlineFirst 03/23/2010.

References

A Genome-Wide Association Study of Prognosis in Breast Cancer

Elizabeth M. Azzato, Paul D.P. Pharoah, Patricia Harrington, et al.


Updated version

Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-10-0085

Cited articles

This article cites 18 articles, 5 of which you can access for free at:
http://cebp.aacrjournals.org/content/19/4/1140.full.html#ref-list-1

Citing articles

This article has been cited by 13 HighWire-hosted articles. Access the articles at:
/content/19/4/1140.full.html#related-urls

E-mail alerts

Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.