Association of Cancer Susceptibility Variants with Risk of Multiple Primary Cancers: The Population Architecture using Genomics and Epidemiology Study


Abstract

Background: Multiple primary cancers account for approximately 16% of all incident cancers in the United States. Although genome-wide association studies (GWAS) have identified many common genetic variants associated with various cancer sites, no study has examined the association of these genetic variants with risk of multiple primary cancers (MPC).

Methods: As part of the National Human Genome Research Institute (NHGRI) Population Architecture using Genomics and Epidemiology (PAGE) study, we used data from the Multiethnic Cohort (MEC) and Women’s Health Initiative (WHI). Incident MPC (IMPC) cases (n = 1,385) were defined as participants diagnosed with more than one incident cancer after cohort entry. Participants diagnosed with only one incident cancer after cohort entry with follow-up equal to or longer than IMPC cases served as controls (single-index cancer controls; n = 9,626). Fixed-effects meta-analyses of unconditional logistic regression analyses were used to evaluate the associations between 188 cancer risk variants and IMPC risk. To account for multiple comparisons, we used the false-positive report probability (FPRP) to determine statistical significance.

Results: A nicotine dependence–associated and lung cancer variant, CHRNA3 rs578776 [OR, 1.16; 95% confidence interval (CI), 1.05–1.26; P = 0.004], and two breast cancer variants, EMBP1 rs11249433 and TOX3 rs3803662 (OR, 1.16; 95% CI, 1.04–1.28; P = 0.005 and OR, 1.13; 95% CI, 1.03–1.23; P = 0.006), were significantly associated with risk of IMPC. The associations for rs578776 and rs11249433 remained (P < 0.05) after removing subjects who had lung or breast cancers, respectively (P ≤ 0.046). These associations did not show significant heterogeneity by smoking status (P_{heterogeneity} ≥ 0.53).

Conclusions: Our study has identified rs578776 and rs11249433 as risk variants for IMPC. These findings may help to identify genetic regions associated with IMPC risk.

Impact: These findings may help to identify genetic regions associated with IMPC risk.

Cancer Epidemiol Biomarkers Prev; 23(11); 2568–78. © 2014 AACR.
comorbidities, including the occurrence of subsequent primary cancers (3), which is one of the leading causes of death for long-term cancer survivors (>5 years; refs. 4, 5). In 2002, Surveillance Epidemiology and End Results (SEER) data indicated that approximately 16% of all new cancers in the United States were second or higher-order primary cancers (1). Risk factors for developing multiple primary cancers (MPC) include younger age at index (first) cancer, treatment modality for the index cancer, shared etiologic factors (e.g., smoking), and genetic susceptibility (3).

Genome-wide association studies (GWAS) have identified several hundred genetic variants associated with various types of cancers (6). Pleiotropy, defined as when a single variant is associated with more than one phenotype, has been identified for some of these cancer risk loci (7, 8). For instance, variants in the chromosome 8q24 locus have been associated with risk of prostate, breast, colorectal, and other cancers (8) and those near the telomerase reverse transcriptase gene (TERT) have been associated with many cancer sites (9–13), including lung (14–16), estrogen receptor (ER)–negative breast (17), ovary (18), and pancreatic (19) cancers. Given that some cancer risk variants are associated with more than one cancer site, it is possible that they may also be associated with risk of developing MPC. Identification of such risk variants may elucidate common cancer etiologies and pathways, improve our understanding of treatment-related effects or identify populations at risk for developing multiple cancers.

With the exception of familial cancer syndromes, the literature examining genetic variants and risk of MPC is sparse, and is usually limited to risk after a specific index cancer site. Furthermore, prior studies are rarely based on prospective data and there has been no systematic investigation of the relationship between common cancer susceptibility variants and MPC risk. Thus, we used prospective data from the Population Architecture using Genomics and Epidemiology (PAGE) study to investigate the associations of 188 cancer risk variants with risk of incident MPC (IMPC).

Materials and Methods
Study populations
The PAGE study (20) was initiated in 2008 by the National Human Genome Research Institute (NHGRI). Two PAGE studies in collaboration with the NHGRI and the coordinating center (20) participated in this analysis, the Multiethnic Cohort (MEC; ref. 21) and the Women’s Health Initiative (WHI; ref. 22). Informed consent was obtained from all study participants. This investigation was approved by each study’s respective Institutional Review Boards.

Briefly, the MEC was initiated in 1993 to investigate the impact of dietary and environmental factors on major chronic diseases, particularly cancer, in ethnically diverse populations in Hawaii and California (21). The study recruited 96,810 men and 118,441 women of ages 45 to 75 years between 1993 and 1996. MEC subjects were contacted mostly from 1995 to 2001 for blood collection included incident cancers with breast, prostate, or colorectal cancers, as well as a random sample of cohort participants to serve as controls in genomic nested case–control studies (participation rate 72% and 63%, respectively). The median interval between diagnosis and blood draw was 14 months (interquartile range, 10–19 months). From 2001 to 2006, blood was also collected, prospectively, without regard for cancer diagnosis, from willing cohort participants (participation rate 43%). Incident cancers are identified through annual linkage to the Hawaii Tumor Registry, the Los Angeles County Cancer Surveillance Program, and the California Cancer Registry. All three registries are members of SEER (23). At the time of case selection for this analysis, follow-up was complete as to December 31, 2007.

The WHI is a long-term health study that recruited 161,808 postmenopausal women of ages 50 to 79 years between 1992 and 1998 at 40 clinical centers throughout the United States. WHI comprises a clinical trial arm, an observational study arm, and several extension studies, which are studies with continued follow-up of consenting participants from the original two study arms. The details of WHI have been previously described (22, 24), and are available online (25). Blood draw was taken at time of study recruitment on all study participants. During the follow-up period of January 1, 1992, to September 30, 2013, incident cancers were identified among participants in the observational study and clinical trial arm. Potential cancer status was identified from self-reported and/or hospital record data and confirmation of status was determined through review of the participant’s medical records and pathology reports by trained cancer coders.

Case and control definitions
Details regarding the study design and case–control definitions are presented in Supplementary Table S1. All subjects for this analysis were cancer-free at baseline and had to have developed at least one cancer during study follow-up. MPC cases were defined as those who were diagnosed with more than one invasive primary cancer (nonmelanoma skin cancers were excluded) during study follow-up (26), and are referred herein as IMPC cases. Controls were defined as subjects who developed only one invasive primary cancer during study follow-up and are referred to as incident single-index cancer (ISC) controls. To test the association of cancer risk variants with risk of subsequent cancers, cancer-free controls were not used in this analysis. In the MEC, which identifies cancer cases by linkage to the SEER registries, IMPC cases could have a subsequent primary cancer of the same site, provided that the subsequent cancer met the SEER definition for same site MPCs [e.g., a subsequent breast cancer following a first breast cancer defined as primary given that did not result from metastases, (ii) breast cancer of a
different histologic cell type, or (iii) breast cancer 5+ years after initial diagnosis; ref. 26). No second primaries were allowed for prostate cancer, leukemia, and non-Hodgkin lymphoma, also based on the SEER rules. In the WHI, which adjudicated cancer cases by medical record review, it was not possible to systematically determine whether the cancer of the same site was a subsequent primary cancer or metastases. Thus, a repeat diagnosis of the same cancer site during follow-up was not counted as IMPC in WHI. IMPC cases for each study were frequency-matched to all eligible ISC controls by age of index cancer diagnosis (within 5-year intervals), length of follow-up (where controls had follow-up time that was equal to or greater than that of the cases), race, and sex. This analysis included a total of 1,385 IMPC cases and 9,626 ISC controls. In preliminary analyses, we had additionally matched for first cancer site. As the findings were similar, we matched only on age, sex, and race/ethnicity to increase our sample size and adjusted for index cancer in the statistical model.

SNP selection and genotyping
A total of 188 cancer risk variants associated with 18 cancers and nicotine dependence (nicotine dependence SNPs were considered herein as cancer risk variants given that smoking is a major risk factor for several cancer sites; ref. 27) were identified from the NHGRI catalog of GWAS studies (6) as of January 2010, followed by a review of the original reports and the fine mapping literature (12, 13–16, 19, 28–83). The risk allele for each SNP was determined on the basis of the prior literature and was defined as the allele associated with an increased risk of cancer in the first GWAS report.

Genotyping for the 188 cancer variants (175 SNPs in MEC; 143 SNPs with overlap) was performed using the TaqMan OpenArray platform (MEC) and Illumina BeadXpress (WHI). To control for population stratification, a panel of 128 ancestry informative markers (84) were genotyped. Principal components analysis (85) was performed and the main principal components were included in the regression model to adjust for genetic ancestry in each study.

Standard quality-assurance and quality-control measures were applied to ensure genotyping quality (20). Samples and SNPs were included on the basis of call rates (≥90%), concordance of blinded replicates (>98%), and no clear departure from Hardy–Weinberg equilibrium (P > 0.001).

Statistical analysis
For each study, we tested the association between each SNP and risk of IMPC using unconditional logistic regression. SNPs were coded additively, with 0, 1, and 2 referring to the number of risk alleles. Models were adjusted for age at diagnosis for the index cancer (continuous), sex (MEC only), study design (WHI only: clinical trial vs. observational study), the most significant principal components of genetic ancestry [to account for race/ethnicity (five principal components for the MEC and three principal components for WHI)], and index cancer site (17 categories), stage (local, regional, and distant), and diagnosis year (to account for treatment cohort effects). We adjusted for index cancer site and stage to account for differences in cancer survival.

The regression estimates were combined across studies using inverse-variance weighted, fixed-effect meta-analysis as implemented in METAL (http://www.sph.umich.edu/csg/abecasis/Metal/). We calculated the cross-study and cross-race heterogeneity P values based on Cochran Q statistic. Stratified analysis by race/ethnicity was conducted for SNP associations with P < 0.05. To account for the multiple testing of 188 SNPs, we used the false-positive report probability (FPRP) introduced by Wacholder and colleagues (86). We set a stringent FPRP threshold of 0.20 and assigned a prior probability range of 0.01 to 0.10 to detect an OR of 1.2 or 0.83. Variants that were found to be significantly associated with IMPC after correction for multiple comparisons were further investigated in sensitivity analyses in which subjects with the cancer site (either as index cancer or IMPC) corresponding to the variant’s first known GWAS association were removed, as well as after adjusting for or stratifying by smoking status, obtained by self-report at baseline.

Results
The main characteristics of the 1,385 IMPC cases and 9,626 ISC controls are presented in Table 1. Subjects were of six different ethnic/racial populations: European American, African American, Hispanic, Asian, Pacific Islander, and Native American. Forty-three percent of subjects were European American and 59% were female (the MEC ascertained both sexes, while WHI ascertained only women). The most common index cancer sites were prostate and breast for the MEC and colorectal and breast for WHI. In both studies, among all IMPC cases, breast cancer was the most common incident second cancer (for the MEC, n = 164 and for WHI, n = 133). The distribution of index and second (subsequent) incident cancers by study is presented in Table 2 and Supplementary Table S2 (which includes latency intervals between cancer sites). The most common index and subsequent cancer combination in the MEC study was an index breast and subsequent breast cancers and in the WHI study index endometrial and subsequent breast cancers. Less common cancers were collapsed as “other”; the enumeration of these cancer sites can be found in Supplementary Table S3 (WHI) and Supplementary Table S4 (MEC). There were 115 MEC IMPC cases and 15 WHI IMPC cases with three or more cancers (Supplementary Table S5).

We examined a total of 188 cancer risk variants (Table 3 for associations with P < 0.05 and Supplementary Table S6 for associations with P ≥ 0.05). Prostate and breast cancer risk variants represented the greatest proportions of variants (28% and 15%, respectively). Among the 188 SNPs, 12 were found to be associated with risk of IMPC at P < 0.05 (Table 3). These 12 associations were with three
prostate cancer variants, three lung cancer or nicotine dependence variants, two breast cancer variants, one pancreatic cancer variant, one melanoma variant, one thyroid variant, and one colorectal variant. These associations were not found to be heterogeneous by study or race/ethnicity (\(P/C21\) 0.05; Table 3 and Supplementary Table S7).

After correction for multiple comparisons, we found that the associations of \(CHRNA3\) rs578776, \(EMBP1\) rs11249433, and \(TOX3\) rs3803662 with MPC remained statistically significant at a FPRP threshold of 0.20 and a prior probability of 0.01 (Table 4). The \(CHRNA3\) rs578776 risk allele “C” previously associated with nicotine

| Table 1. Study characteristics for IMPC cases and first index cancer controls |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
|                             | MEC (\(n = 965\)) | Controls (\(n = 6,969\)) | WHI (\(n = 420\)) | Controls (\(n = 2,657\)) |
| Age of index cancer; mean (SD) | 68.3 (7.8) | 68.0 (8.3) | 70.1 (7.0) | 69.7 (9.0) |
| Age of second cancer; mean (SD) | 72.6 (8.0) | NA | 74.2 (7.3) | NA |
| Follow-up time, y; mean (SD) | 9.4 (3.3) | 13.7 (1.9) | 9.6 (3.7) | 12.7 (2.3) |
| Sex; n (%) | | | | |
| Male | 549 (56.9) | 3,930 (56.4) | 420 (100) | 2,657 (100) |
| Female | 416 (43.1) | 3,039 (43.6) | | |
| Race/ethnicity; n (%) | | | | |
| European American | 254 (26.3) | 1,534 (22.0) | 372 (88.6) | 2,088 (78.6) |
| African American | 235 (24.4) | 1,419 (20.4) | 37 (8.8) | 412 (15.5) |
| Hispanic | 194 (20.1) | 1,581 (22.7) | 6 (1.4) | 96 (3.6) |
| Asian | 232 (24.0) | 2,045 (29.3) | 4 (0.95) | 58 (2.2) |
| Pacific Islander | 50 (5.2) | 390 (5.6) | 1 (0.24) | 3 (0.11) |
| Indian/Native American | NA | NA | 1 (0.24) | 3 (0.11) |
| Smoking status; n (%) | | | | |
| Never | 352 (36.9) | 2,843 (41.4) | 174 (41.4) | 1,292 (48.6) |
| Former | 421 (44.1) | 3,064 (44.6) | 200 (47.6) | 1,123 (42.3) |
| Current | 182 (19.1) | 959 (14.0) | 42 (10.0) | 212 (8.0) |
| Missing, n | 10 | 103 | 4 | 30 |
| Pack-years\(\^a\); mean (SD) | 21.9 (17.0) | 18.0 (15.7) | 15.3 (22.0) | 11.0 (19.1) |
| Missing, n | 31 | 260 | 14 | 90 |
| Index cancer site; n (%) | | | | |
| Breast | 265 (27.5) | 1,970 (28.3) | 82 (19.5) | 906 (34.1) |
| Prostate | 278 (28.8) | 2,906 (41.7) | NA | NA |
| Colorectal | 164 (17.0) | 987 (14.2) | 67 (16.0) | 433 (16.3) |
| Lung | 24 (2.5) | 175 (2.5) | 43 (10.2) | 291 (11.0) |
| Other\(\^b\) | 234 (24.2) | 931 (13.4) | 228 (54.3) | 1,027 (38.7) |
| Secondary cancer site; n (%) | | | | |
| Breast | 164 (17.0) | NA | 133 (31.7) | NA |
| Prostate | 111 (11.5) | NA | NA | NA |
| Colorectal | 158 (16.4) | NA | 49 (11.7) | NA |
| Lung | 120 (12.4) | NA | 65 (15.5) | NA |
| Other\(\^b\) | 412 (42.7) | NA | 173 (41.2) | NA |
| Stage of index cancer; n (%) | | | | |
| Localized | 683 (70.8) | 5,153 (73.9) | 288 (68.6) | 1,651 (61.1) |
| Regional | 215 (22.3) | 1,346 (19.3) | 86 (20.1) | 640 (24.1) |
| Distant | 67 (6.9) | 470 (6.7) | 46 (11.0) | 366 (13.8) |
| Diagnosis year for index cancer, mean (SD) | 1998 (3.4) | 2000 (3.7) | 2002 (3.5) | 2003 (3.7) |
| Diagnosis year for second cancer, mean (SD) | 2002 (3.2) | NA | 2006 (3.7) | NA |

Abbreviation: NA, not applicable.
\(\^a\)Pack-years for ever smokers only.
\(\^b\)Counts for Asian and Pacific Islanders.
\(\^c\)The four most common other cancer sites include bladder, endometrium, leukemia, melanoma, and non-Hodgkin lymphoma; enumeration for these sites can be found in Table 2.
dependence and lung cancer was associated with an increased risk of IMPC [OR, 1.15; 95% confidence interval (CI) = 1.05–1.26; \( P = 0.004; \) Table 2]. The risk allele “C” for the breast cancer variant EMBP1 rs11249433 was also associated with an increased risk of IMPC (OR, 1.16; 95% CI, 1.04–1.28; \( P = 0.005\)). The risk allele “T” for the breast cancer variant TOX3 rs3803662, located at 16q12.1, was also associated with an increased risk of IMPC (OR, 1.13; 95% CI, 1.03–1.23; \( P = 0.006\)).

When removing from the analysis all subjects with lung cancer for the CHRNA3 variant and with breast cancer for the EMBP1 and TOX3 variants, either as index cancer or IMPC, the associations for CHRNA3 rs578776 and EMBP1 rs11249433 remained significant (\( P < 0.046\); Table 5). The association for rs3803662 had a similar OR as its main effects; however, the association did not quite reach statistically significant (OR, 1.13; \( P = 0.005\) vs. OR, 1.11; \( P = 0.063\), respectively).

Because smoking is a risk factor for many cancers, we additionally adjusted our models for smoking status and pack-years and found similar associations (data not shown). In our analyses stratified by smoking status, no heterogeneity in the effects of the genetic variants was detected between never and ever smokers (\( P_{\text{heterogeneity}} > 0.53\); Table 5). Because of a limitation in sample size, we were unable to restrict this analysis to cancers that are unrelated to smoking.

**Discussion**

To our knowledge, this is the first study to examine the association of cancer risk variants with risk of IMPC. We tested 188 established cancer risk variants among 1,385 IMPC cases and 9,626 ISC controls and found that the lung cancer/nicotine dependence risk allele “C” of CHRNA3 rs578776 and the breast cancer risk alleles “C” of EMBP1 rs11249433 and “T” of TOX3 of rs3803662 were associated with an increased risk of IMPC. The associations for rs578776 and rs11249433 remained after accounting for multiple hypothesis testing, were not restricted to participants with the cancer previously associated with those variants, and were consistent with regard to smoking status.

The 15q25.1 region includes the CHRNA5–CHRNA3–CHRNB4 cluster of cholinergic nicotine receptor subunit genes. Variants within this region have been associated with increased risk of lung cancer primarily due to their relationships with smoking behavior and nicotine dependence (14, 80, 81, 87). In addition, risk variants in the CHRNA5–CHRNA3–CHRNB4 region have been positively associated with chronic obstructive pulmonary disease (88), serum albumin levels (89), pulmonary function (90), and childhood obesity–related traits in Hispanics (91). In our study, we found that both rs578776 and rs8042374, which are in high linkage disequilibrium (LD) in whites and Asians [Utah residents with ancestry from northern and western Europe (CEU): \( r^2 = 0.83\); Yoruba in Ibadan, Nigeria (YRI): \( r^2 = 0.16\); Japanese in Tokyo, Japan and Han
<table>
<thead>
<tr>
<th>SNP</th>
<th>First GWAS citation</th>
<th>First GWAS site</th>
<th>Region</th>
<th>Position</th>
<th>Mapped gene</th>
<th>Risk allele/ ref. allele</th>
<th>Meta-OR (95% CI)</th>
<th>$P$</th>
<th>Direction</th>
<th>$P_{het}$ for study</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs578776</td>
<td>Saccone et al., 2009 (83)</td>
<td>Nicotine dependence</td>
<td>15q25.1</td>
<td>76,675,455</td>
<td>CHRNA3</td>
<td>1,377</td>
<td>9,559 C/T</td>
<td>0.49</td>
<td>1.15</td>
<td>(1.05–1.26) 0.004 +</td>
</tr>
<tr>
<td>rs11249433</td>
<td>Thomas et al., 2009 (68)</td>
<td>Breast cancer</td>
<td>1p11.2</td>
<td>120,982,136</td>
<td>EMBP1</td>
<td>1,347</td>
<td>9,406 C/T</td>
<td>0.26</td>
<td>1.16</td>
<td>(1.04–1.28) 0.005 +</td>
</tr>
<tr>
<td>rs3803662</td>
<td>Thomas et al., 2009 (68)</td>
<td>Breast cancer</td>
<td>16q12.1</td>
<td>51,143,842</td>
<td>TOX3</td>
<td>1,377</td>
<td>9,577 T/C</td>
<td>0.40</td>
<td>1.13</td>
<td>(1.03–1.23) 0.006 +</td>
</tr>
<tr>
<td>rs4975616</td>
<td>Broderick et al., 2009 (15)</td>
<td>Lung cancer</td>
<td>5p15.33</td>
<td>1,368,660</td>
<td>Intergenic</td>
<td>1,378</td>
<td>9,576 A/G</td>
<td>0.61</td>
<td>0.80</td>
<td>(0.82–0.98) 0.014 −</td>
</tr>
<tr>
<td>rs8042374</td>
<td>Wang et al., 2008 (14)</td>
<td>Lung cancer</td>
<td>15q25.1</td>
<td>76,695,087</td>
<td>CHRNA3</td>
<td>1,368</td>
<td>9,550 A/G</td>
<td>0.57</td>
<td>1.12</td>
<td>(1.02–1.24) 0.015 +</td>
</tr>
<tr>
<td>rs4857841</td>
<td>Gudmundsson et al., 2009 (46)</td>
<td>Prostate cancer</td>
<td>9q21.3</td>
<td>129,529,333</td>
<td>EEFSEC</td>
<td>1,376</td>
<td>9,534 A/G</td>
<td>0.44</td>
<td>0.90</td>
<td>(0.82–0.98) 0.020 −</td>
</tr>
<tr>
<td>rs4785763</td>
<td>Bishop et al., 2009 (31)</td>
<td>Melanoma</td>
<td>16q24.3</td>
<td>88,594,437</td>
<td>AFG3L1P</td>
<td>1,370</td>
<td>9,551 A/C</td>
<td>0.28</td>
<td>1.11</td>
<td>(1.01–1.22) 0.023 +</td>
</tr>
<tr>
<td>rs966513</td>
<td>Gudmundsson et al., 2009 (46)</td>
<td>Thyroid cancer</td>
<td>9q22.33</td>
<td>99,595,930</td>
<td>Intergenic</td>
<td>1,377</td>
<td>9,562 A/G</td>
<td>0.24</td>
<td>0.89</td>
<td>(0.80–0.99) 0.026 −</td>
</tr>
<tr>
<td>rs5945619</td>
<td>Eeles et al., 2008 (42)</td>
<td>Prostate cancer</td>
<td>Xp11.22</td>
<td>51,258,412</td>
<td>Intergenic</td>
<td>1,378</td>
<td>9,566 C/T</td>
<td>0.27</td>
<td>1.09</td>
<td>(1.01–1.18) 0.035 +</td>
</tr>
<tr>
<td>rs3790844</td>
<td>Petersen et al., 2010 (19)</td>
<td>Pancreatic cancer</td>
<td>1q32.1</td>
<td>198,274,055</td>
<td>NR5A2</td>
<td>420</td>
<td>2,657 T/C</td>
<td>0.77</td>
<td>1.23</td>
<td>(1.01–1.49) 0.035 +</td>
</tr>
<tr>
<td>rs10096908</td>
<td>Al Olama et al., 2009 (29)</td>
<td>Prostate cancer</td>
<td>8q24.21</td>
<td>128,041,119</td>
<td>Intergenic</td>
<td>957</td>
<td>6,904 T/C</td>
<td>0.74</td>
<td>1.13</td>
<td>(1.01–1.27) 0.40 x</td>
</tr>
<tr>
<td>rs4939827</td>
<td>Teneva et al., 2008 (67)</td>
<td>Colorectal cancer</td>
<td>18q21.1</td>
<td>44,707,461</td>
<td>SMAD7</td>
<td>1,374</td>
<td>9,532 T/C</td>
<td>0.39</td>
<td>1.09</td>
<td>(1.00–1.19) 0.047 +</td>
</tr>
</tbody>
</table>

Abbreviation: RAF, risk allele frequency.

*Models were adjusted for age at diagnosis for the index cancer (continuous), sex (MEC only), study design (WHI only: clinical trial vs. observational study), the most significant principal components.

+$+$, positive association; $-$, negative association; $\times$, not tested.
Chinese in Beijing, China (JPT+CHB; $r^2 = 0.72$), were positively associated with risk of IMPCs when compared with those with only one incident cancer. Only the association with rs578776 remained significant after correction for multiple testing. Smoking is an established risk factor for at least 10 different cancer sites (27); therefore, a reasonable explanation for the association of these variants with IMPC would be the result of their known association with smoking. However, when stratifying by smoking status, we found that the association was somewhat stronger, albeit not significantly ($P_{\text{heterogeneity}} > 0.47$), in never smokers (OR, 1.21; $P = 0.01$) than in ever smokers (OR, 1.13; $P = 0.05$). This association in never smokers is unexpected and could be due to chance. Further study of this risk variant with IMPC in never smokers is warranted.

The rs11249433 variant is located in the pericentromeric region at 1p11.2 and within the embigin pseudogene, EMBPI. This SNP maps near many other pseudogenes in a SNP desert region. EMBPI rs11249433 was originally found to be associated with breast cancer ($P = 7 \times 10^{-10}$, ref. 68). Additional studies found that rs11249433 was

### Table 4. FPRP values for the 10 SNPs associated with IMPC risk

<table>
<thead>
<tr>
<th>SNP</th>
<th>First GWAS site</th>
<th>Meta-OR (95% CI)$^a$</th>
<th>Power$^b$</th>
<th>Reported $P$</th>
<th>FPRP based on prior:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>rs578776</td>
<td>Nicotine dependence</td>
<td>1.15 (1.05–1.26)</td>
<td>0.82</td>
<td>0.004</td>
<td>0.03</td>
</tr>
<tr>
<td>rs11249433</td>
<td>Breast cancer</td>
<td>1.16 (1.04–1.28)</td>
<td>0.75</td>
<td>0.005</td>
<td>0.04</td>
</tr>
<tr>
<td>rs3803662</td>
<td>Breast cancer</td>
<td>1.13 (1.03–1.23)</td>
<td>0.92</td>
<td>0.006</td>
<td>0.44</td>
</tr>
<tr>
<td>rs4975616</td>
<td>Lung cancer</td>
<td>0.89 (0.82–0.98)</td>
<td>0.93</td>
<td>0.014</td>
<td>0.15</td>
</tr>
<tr>
<td>rs8042374</td>
<td>Lung cancer</td>
<td>1.12 (1.02–1.24)</td>
<td>0.91</td>
<td>0.015</td>
<td>0.12</td>
</tr>
<tr>
<td>rs4857841</td>
<td>Prostate cancer</td>
<td>0.90 (0.82–0.98)</td>
<td>0.96</td>
<td>0.020</td>
<td>0.13</td>
</tr>
<tr>
<td>rs4785763</td>
<td>Melanoma</td>
<td>1.11 (1.01–1.22)</td>
<td>0.95</td>
<td>0.023</td>
<td>0.22</td>
</tr>
<tr>
<td>rs965513</td>
<td>Thyroid cancer</td>
<td>0.89 (0.80–0.99)</td>
<td>0.89</td>
<td>0.026</td>
<td>0.24</td>
</tr>
<tr>
<td>rs9545619</td>
<td>Prostate cancer</td>
<td>1.09 (1.01–1.18)</td>
<td>0.99</td>
<td>0.035</td>
<td>0.21</td>
</tr>
<tr>
<td>rs3790844</td>
<td>Pancreatic cancer</td>
<td>1.23 (1.01–1.49)</td>
<td>0.40</td>
<td>0.035</td>
<td>0.44</td>
</tr>
<tr>
<td>rs10086708</td>
<td>Prostate cancer</td>
<td>1.13 (1.01–1.27)</td>
<td>0.84</td>
<td>0.040</td>
<td>0.30</td>
</tr>
<tr>
<td>rs4938927</td>
<td>Colorectal cancer</td>
<td>1.09 (1.00–1.19)</td>
<td>0.98</td>
<td>0.047</td>
<td>0.33</td>
</tr>
</tbody>
</table>

NOTE: The values in bold indicate that the associations remained significant at an FPRP threshold of 0.20 and a prior probability range of 0.1 through 0.001 of detecting an OR of 1.2 or 0.83.

$^a$ORs from Table 2, models were adjusted for age at diagnosis for the index cancer (continuous), sex (MEC only), study design (WHI only: clinical trial vs. observational study), the most significant principal components of genetic ancestry (five principal components for the MEC and three principal components for WHI), index cancer site, stage, and diagnosis year (to account for treatment cohort effects).

$^b$Statistical power calculated using the recessive model, except where noted, is the power to detect an OR of 1.2 or 0.83 at a level of 0.05.

### Table 5. Associations of two cancer risk variants with IMPC risk, after subjects with the cancer site corresponding to the known association have been removed, or after stratifying by smoking status

<table>
<thead>
<tr>
<th>SNP</th>
<th>First GWAS site</th>
<th>Meta-OR (95% CI)$^a$</th>
<th>Power$^b$</th>
<th>Reported $P$</th>
<th>FPRP based on prior:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>Cancer site removed$^a$</td>
<td>Breast cancer SNP rs11249433</td>
<td>1.15 (1.00–1.31)</td>
<td>0.046</td>
<td>841</td>
<td>6,712</td>
</tr>
<tr>
<td>Stratify by smoking$^b$</td>
<td>Cancer site removed$^a$</td>
<td>822</td>
<td>6,596</td>
<td>0.046</td>
<td>841</td>
</tr>
<tr>
<td>Never smokers</td>
<td>509</td>
<td>4,037</td>
<td>1.12 (1.04–1.32)</td>
<td>0.20</td>
<td>523</td>
</tr>
<tr>
<td>Ever smokers</td>
<td>736</td>
<td>5,034</td>
<td>1.19 (1.04–1.36)</td>
<td>0.01</td>
<td>809</td>
</tr>
<tr>
<td>$P_{\text{heterogeneity}}$ by smoking status</td>
<td>0.53</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$Subjects with the cancer site corresponding to the known association with cancer have been removed.

$^b$Models were adjusted for age at diagnosis for the index cancer (continuous), sex (MEC only), study design (WHI only: clinical trial vs. observational study), the most significant principal components of genetic ancestry.

$^c$Additionally adjusted for pack-years.
specifically associated with ER-positive breast cancer (68, 92) and lower-grade breast tumors (92), two characteristics that are associated with improved breast cancer survival (93). It is possible that the association for rs11249433 with increased risk of MPC may be related to survivorship such that a longer survival after breast cancer diagnosis may lead to a higher likelihood of developing a subsequent primary cancer. When removing subjects with breast cancer as the index cancer or an IMPC from our analysis, the association of rs11249433 with IMPC persisted, although not as significant, suggesting that this association was not exclusively due to a better breast cancer survival. This SNP may play a role in increased overall cancer survival or risk of cancer development. The closest located genes to rs11249433 are a transmembrane coding gene, NOTCH2, and highly paralogous low-affinity Fc gamma receptor family 1B gene, FCGR1B. SNPs in NOTCH2 have been associated with type II diabetes (94), a condition that has been associated with an increased risk of breast, colorectal, pancreatic, liver, urinary tract, and endometrial cancers (95, 96). In addition, a study that investigated the expression of five genes within 1 Mb of rs11249433 found an increased expression of NOTCH2 in patients with ER-positive breast cancer without TP53 mutations compared with those with TP53 mutations (97). NOTCH2 receptors are involved in regulation of cell communication, proliferation, differentiation, and death (98), processes that are all influential in cancer development.

The rs3803662 variant is located at 16q12.1 in the Cancer Susceptibility Candidate 16 gene, CASC16, which is an RNA gene. TOX3 and LOC643714 are other genes located close to this polymorphism. This variant has been previously associated with breast cancer (P = 6 × 10^{-13}, ref. 64), with slightly stronger associations in ER-positive breast cancer than ER-negative disease (99). A prior study found that the risk allele for rs3803662 is correlated with lower mRNA expression of TOX3 in ER-positive tumors (100). Increased TOX3 mRNA expression has been previously found to be predictive of breast cancer metastases (101) and lower overall survival among patients with breast cancer (100). It is possible that the association for rs3803662 with increased risk of MPC may be related to survival from breast cancer. When removing subjects with breast cancer, the association was not quite statistically significant, suggesting that this association may be due to the risk variants prior association with breast cancer.

This study had a number of strengths, including the prospective design that allowed us to focus on incident cancers and minimize survival bias, while investigating MPC risk in an adult population. We also had a long follow-up period to assess subsequent cancers (MEC, 14 years and WHI, 19 years), a large sample size, and well-characterized study populations enabling adjustment for multiple potential confounders. Study limitations include limited power to detect effects for rare SNPs and the lack of detailed information on treatment, especially, on radiotherapy, a known risk factor for multiple primaries (102).

In the MEC, where information on first course of cancer therapy was available, similar results were observed after adjusting for radiation treatment (yes/no). In addition, our meta-analyzed findings did not change when adjusting for index cancer stage, suggesting that our findings are not strongly dependent on the treatment received, because treatment is relatively well standardized by stage. Our study oversampled the more common cancer sites as a result of the blood collection design in the earlier years of the MEC. Also, for WHI, the study undersampled breast and colorectal cancer cases to be genotyped for the PAGE analysis. Therefore, the more common cancer sites, with the exception of lung cancer, may have been overrepresented in the MEC and endometrial cancer was overrepresented in WHI. In addition, cases with subsequent primary cancers of the same site were not available in the WHI study and were, thus, underrepresented in that study.

In conclusion, our findings, if reproduced, may contribute toward identifying common cancer etiologic pathways, common treatment-related effects, or populations at risk for developing more than one cancer. We found the cancer risk variant rs578776 in CHRNA3 to be associated with occurrence of MPCs. Our analysis suggests that the associations for TOX3 rs3803662 may have been driven by the variants known association with breast cancer. However, the association for EMBP1 rs11249433 and CHRNA3 rs578776 appears to be independent of their known cancer association. These findings should be confirmed in other study populations.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Disclaimer
The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Authors’ Contributions
Conception and design: S.L. Park, F.R. Schumacher, U. Peters, L.R. Wilkens, C. Kooperberg, L. Le Marchand
Development of methodology: S.L. Park, C. Kooperberg
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): F.R. Schumacher, U. Peters, G. Heiss, D.C. Crawford, L.R. Wilkens, W.S. Bush, C. Kooperberg, L. Le Marchand
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.L. Park, Y. Lin, R.J. Goodloe, L. Dumitrescu, J.H. Fowke, L. Hou, Y. Han, L.R. Wilkens, W.S. Bush, I. Cheng, L. Le Marchand
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.L. Park, C.P. Caberto, S.-A. Love, T.C. Matise, L.A. Hindorff, E. Deelman, K.E. North, G. Heiss, L. Le Marchand
Study supervision: K.E. North, G. Heiss, D.C. Crawford, C. Kooperberg, L. Le Marchand

Acknowledgments
The authors thank the MEC and 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: https://www.whi.org/researchers/SitePages/WHI%20Investigators.aspx WHI_investigators...
shortlist.pdf. The PAGE consortium thanks the staff and participants of all PAGE studies for their important contributions. The complete list of PAGE members can be found at http://www.pagestudy.org. The authors gratefully acknowledge the time and efforts of all the study participants.

Grant Support

(a) The PAGE program is funded by the NHGRI supported by U01HG004803 (CALICo; awarded to K.E. North and G. Heiss), U01HG004798 (EAGLE; awarded to D.C. Crawford), U01HG004802 (MEC; awarded to L. Le Marchand), U01HG004790 (WHI; awarded to C. Kooperberg), and U01HG004801 (Coordinating Center; awarded to T.C. Matise), and their respective NHGRI American Recovery and Reinvestment Act of 2009 (ARRA) supplements. (b) The data and materials included in this report result from a collaboration between the following studies: S.L. Park, C.P. Caberto, I. Cheng, L. Le Marchand, C.A. Haiman, L.R. Wilkens, F.R. Schumacher, and Y. Han are affiliated with the MEC study. The MEC study characterization of epidemiologic architecture is funded through the NHGRI PAGE program (U01HG004802 and its NHGRI ARRA supplement). The MEC study is funded through the National Cancer Institute (R37CA54281, R01CA33619, U01CA136792, and U01CA9750). This work is also supported, in part, by NIH grant P30 CA014089 to the Norris Comprehensive Cancer Center (S.L. Park). Funding support for the "Epidemiology of putative genetic variants: The Women’s Health Initiative" study is provided through the NHGRI PAGE program (U01HG004790 and its NHGRI ARRA supplement). The WHI program is funded by the National Heart, Lung, and Blood Institute; NIH; and U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32105-2, 32108-9, 32111-3, 32115, 32118-32119, 32122, 42107-26, 42129-32, and 4421. W.S. Bush, D.C. Crawford, L. Dumitrescu, J.H. Fowke, and R.J. Goodell are affiliated with the EAGLE study, which is supported by NHGRI U01HG004798. G. Heiss, S. Love, and K.E. North are affiliated with the CALICo consortium, which is supported by NHGRI U01HG004803. E. Deelman and T.C. Matise are affiliated with the Coordinating Center, which is supported by the NHGRI U01HG004801. Assistance with phenotype harmonization, SNP selection and annotation, data cleaning, data management, integration and dissemination, and general study coordination was provided by the PAGE Coordinating Center (U01HG004801-01 and its NHGRI ARRA supplement). The National Institutes of Mental Health also contributes to the support for the Coordinating Center. L.A. Hindorff is an epidemiology and program director in the Division of Genomic Medicine at the NHGRI. She is the project scientist for the PAGE study.

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 March 5, 2014; revised July 25, 2014; accepted August 5, 2014; published OnlineFirst August 19, 2014.

References


