

SHORT REPORT

Caution in generalizing known genetic risk markers for breast cancer across all ethnic/racial populations

Fang Chen¹, Daniel O Stram¹, Loïc Le Marchand², Kristine R Monroe¹, Laurence N Kolonel², Brian E Henderson¹ and Christopher A Haiman^{*,1}

Genome-wide association studies (GWAS) have identified common variants associated with breast cancer risk among women of European and Asian ancestries. To assess the generalizability across ethnic/racial populations of a risk score derived from genotyping 12 highly replicated breast cancer GWAS hits, we performed a case-control study (2224 cases and 2827 controls) nested in the Multiethnic Cohort (MEC) study, which was initiated in 1993–1996 and consists of subjects mainly from European-American, African-American, Native Hawaiian, Japanese and Latino populations. When viewed as a summary risk score, the total number of risk alleles carried by women was significantly associated with breast cancer risk overall (OR per allele, 1.09; 95% CI, 1.06–1.12; $P=2.0\times 10^{-10}$) and in all populations except African-Americans, in which no significant association was observed (OR, 1.03; 95% CI, 0.98–1.08). In aggregate, the number of risk alleles is strongly associated with breast cancer risk in all populations studied except African-Americans. These results emphasize the need for large-scale association studies of multiple racial/ethnic groups for discovery and characterization of risk alleles relevant to all populations in the United States.

European Journal of Human Genetics (2011) 19, 243–245; doi:10.1038/ejhg.2010.185; published online 24 November 2010

Keywords: breast cancer; risk markers; summary score; generalizability; diverse populations

INTRODUCTION

Genome-wide association studies (GWAS) of breast cancer have substantiated a role for common low-risk alleles in disease susceptibility.^{1–6} To date, discovery and replication efforts have been limited primarily to populations of European ancestry, which traditionally have been the focus of most genetic studies of cancer. Prediction of individual risk on the basis of multiple risk markers is one potential utility of this new genetic information, although it has been persuasively argued^{7–10} that many more common low penetrance genetic markers would need to be identified before they would have, individually or as a group, any public health utility. Nevertheless, private companies have already begun to market genetic tests for the currently known risk markers, such as: deCODE BreastCancer (deCODE, Reykjavik, Iceland) for women of European ancestry before the causal alleles underlying the marker associations have been identified. Whether these tests have any relation to risk at all in non-White populations, which make up a third of the US population, is not known.

Associations with cancer risk alleles may not be consistent across populations for a number of reasons,¹¹ including differences by race/ethnicity of the linkage disequilibrium patterns relating a risk marker to the causal variant(s), and/or context dependency of the association resulting from genetic and environmental modifiers that vary in frequency across populations. Studies in multiple populations^{12–14} are needed to examine the generalizability of these markers before their potential for public health utility can be applied to populations of non-European ancestry. Here, we report on the association of 12 validated risk alleles identified in breast cancer GWAS conducted

primarily in populations of European ancestry, among European-American, African-American, Native Hawaiian, Japanese-American and Latino breast cancer cases and controls from the Multiethnic Cohort (MEC) study.¹⁵

MATERIALS AND METHODS

Study population: the MEC

The MEC study is a prospective cohort study initiated between 1993 and 1996. The study consists of 215 251 adult men and women living in Hawaii and California (mainly Los Angeles County) mainly from the following populations: European-American, African-American, Native Hawaiian, Japanese-American and Latino. Drivers' license files were used as a primary source to identify the study subjects. Participants entered the cohort study by completing and returning a self-administered questionnaire that asked information about general demographic characteristics as well as known breast cancer risk factors. Cases were identified through cohort linkage to population-based cancer surveillance, epidemiology and end results (SEER) registries in California and Hawaii.¹⁵ Through December 31, 2005, the breast cancer case-control study nested in the MEC assembled for genetic studies included 2224 cases and 2827 controls frequency-matched on race/ethnicity and age. For this study, we included additional African-American controls to allow for more precise risk estimation. The median ages of cases and controls were 66 and 65 years, respectively, and ranged from 44 to 87. This study was approved by the Institutional Review Boards at the University of Southern California and at the University of Hawaii.

Laboratory assays

We genotyped 12 SNPs from GWAS of breast cancer.^{1–6} We also tested one additional variant in *FGFR2* that was revealed by fine-mapping, which the

¹Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA; ²Epidemiology Program, Cancer Research Center of Hawaii, University of Hawaii, Honolulu, HI, USA

*Correspondence: Dr CA Haiman, Harlyne Norris Research Tower, 1450 Biggy Street, Room 1504, Los Angeles, CA 90033, USA. Tel: +1 323 442 7755; Fax: +1 323 442 7749; E-mail: haiman@usc.edu

Received 3 June 2010; revised 22 September 2010; accepted 7 October 2010; published online 24 November 2010

MEC also participated in identifying (African-Americans only).¹⁶ Genotyping was performed using the TaqMan allelic discrimination assay.¹⁷ We substituted rs10483813 for rs999737 (14q24) as genotyping of the latter failed; the two SNPs have perfect correlation in European-Americans ($r^2=1$). The overall genotyping call rate ranged from 94.6 to 100.0% (average, 97.9%) for the 13 variants. For blinded duplicates, the mismatch rate was <2% for all 13 SNPs (average <1%). Hardy–Weinberg equilibrium (HWE) testing was conducted for each variant in each population using a 1-df χ^2 -test, and all 13 variants were consistent with HWE using a criterion of $P>0.01$ in controls (Supplementary Table 1).

Statistical analysis

In each of the five racial/ethnic groups constituting the MEC, we examined the distribution and breast cancer risk associated with an unweighted summary score, taken as the number of risk alleles for 12 variants (1p11, rs11249433; 2q35, rs13387042; 3p24, rs4973768; 5p12, rs10941679; 5q11, rs889312; 6q25, rs2046210; 8q24, rs13281615; 10q26, rs2981582; 11p15, rs3817198; 14q24, rs10483813; 16q12, rs3803662; 17q23, rs6504950) (Supplementary Table 2) in order to determine their combined contribution to breast cancer risk. Analysis was conducted on 2171 cases and 2795 controls as individuals missing greater than or equal to four SNP genotypes were excluded (53 (2.4%) cases and 32 (1.1%) controls). Missing genotypes of individuals were given the mean score for that locus within each population. Odds ratios were estimated for this risk score, which ranged from 4 to 18 risk alleles per individual, with a median of 11, over all 5 racial/ethnic groups. Odds ratios were adjusted for age (quartiles) and race. As ancestry may differ by case-control status, SNPs may be associated with risk simply because they vary in frequency across racial/ethnic groups. Although we adjust for self-reported ethnicity, several of the populations we consider here are known to be admixed between two or more ancestral groups. We used principal components analysis¹⁸ to control for hidden population stratification (including admixture) that could otherwise cause confounding of unreported ethnicity (or ethnic mixture) with SNP effects. Specifically, we computed the first 10 eigenvectors for principal components analysis using a panel of >1300 SNPs from previous studies not linked to the 12 markers of interest here.¹⁹ These were included as adjustment variables in all models. All statistical analysis was performed in a SAS 9.1 package, SAS Institute Inc., Cary, NC, USA.

RESULTS

Using the unweighted summary score, we observed a highly significant association with breast cancer risk in an ethnic-pooled analysis (per allele: OR, 1.09; 95% CI, 1.06–1.12; $P=2.0\times 10^{-10}$) (Table 1), with women in the upper quintile having a 1.6-fold greater risk than women in the bottom quintile (>12 alleles *vs* <9 alleles; 95% CI, 1.32–1.97; $P=3.0\times 10^{-6}$) (Supplementary Table 3) and women in the highest decile having a 2.0-fold greater risk of breast cancer, compared with those in the bottom decile (>13 alleles *vs* <8 alleles; 95% CI, 1.52–2.73; $P=2.3\times 10^{-6}$).

However, significant racial/ethnic heterogeneity was noted ($P=0.030$, 4-df test). Specifically, the summary score variable was found to be positively and significantly associated with breast cancer risk ($P\leq 3.9\times 10^{-3}$) with effects per allele of ≥ 1.10 in all populations,

except in African-Americans (per allele: OR=1.03, 95% CI=0.98–1.08, $P=0.23$; Table 1). The apparent lack of an association between breast cancer risk either with this aggregate allele count variable, and with many of the individual SNPs (Supplementary Table 4), in African-Americans suggests that few of these variants are likely to be markers of risk in the African-American population.

Given that several of the validated risk alleles have been more strongly associated with ER-positive disease,^{2,3,5,20} we tested for heterogeneity of the risk score by ER status in each population. As is expected, the score was more informative for ER-positive disease than ER-negative disease (Supplementary Table 5). We observed the same pattern of association in ER-positive disease as in the overall pooled analysis, with the summary risk score being significantly associated with disease risk in most populations, and no significant association observed in African-Americans (Supplementary Table 5).

We tested for dominant and recessive effects of single SNPs and observed no significant evidence of a better model fit when the genotypes for each SNP were modeled in combination with the summary risk score (Supplementary Table 6). Odds ratios estimated over the range of observed allele counts were also inspected and were found to be very consistent with the assumption of linear allelic effect (Supplementary Table 3). We also tested for pair-wise gene by gene interactions and observed seven nominally significant interactions; however, none remained statistically significant after Bonferroni correction for multiple comparisons. We also constructed a second summary score, weighting each SNP by its published log OR (Supplementary Table 2). This score measure was highly correlated with the unweighted score ($r=0.88$), and was not superior to the unweighted score when included in the same model.

DISCUSSION

An ultimate public health goal of mapping risk alleles is to predict individual risk so that we can identify those at greater risk, among whom targeted intervention and preventive measures may be applied. An understanding of the polygenic component to breast cancer risk would undoubtedly add significantly to risk prediction,^{7,8} and to the efficacy of population-based programs for prevention and early detection. Nevertheless, as noted by Gail,⁹ a much larger number of modestly penetrant risk variants may be needed to make a significant impact on the problem of breast cancer risk prediction. Replicating aggregate allele counts, or other summary variables, thus, is much more important to this problem than the replication of any one specific risk marker. Our sample sizes were too small to fully replicate, in any single racial/ethnic group, the modest risks that each of the 12 validated markers have shown in Whites (Supplementary Table 2 and 4); however, we did have very good power to detect the effect of the aggregate variable in any specific racial/ethnic group assuming homogeneity of effect by ethnic group.

Table 1 The summary associations of validated breast cancer risk variants in diverse populations

	European-Americans 534/541 ^a	African-Americans 539/1022	Native Hawaiians 146/287	Japanese 561/550	Latinos 391/395	Pooled 2171/2795	P_{int}^b
Per allele OR (95% CI) ^c	1.11 (1.05–1.18)	1.03 (0.98–1.08)	1.21 (1.09–1.35)	1.10 (1.03–1.17)	1.12 (1.05–1.20)	1.09 (1.06–1.12)	—
P-value	1.6×10^{-4}	0.23	4.2×10^{-4}	3.9×10^{-3}	8.5×10^{-4}	2.0×10^{-10}	0.030

Abbreviations: OR, odds ratio; CI, confidence interval.

^aNumber of cases/controls.

^bP-value for interaction between the summary score and ethnic group (4-df).

^cOR adjusted for age(quartiles), the first 10 eigenvectors from principal components analysis, and race (in pooled analysis).

Possible explanations for the lack of association with the aggregate variable in African-Americans is that the majority of true risk alleles underlying the marker associations are rare in African-Americans and/or linkage disequilibrium does not extend as far in persons of African ancestry. Both possibilities emphasize the need to conduct full-scale high-density association studies to identify racial/ethnic specific risk markers or to further refine the association signals in the regions containing these risk alleles in racial/ethnic groups. For example, fine-mapping of *FGFR2* has revealed a stronger marker of risk in African-Americans (rs2981578).¹⁶ This marker made some improvement to risk prediction with the aggregate score in this population (per allele: OR, 1.05; 95% CI, 1.00–1.10, $P=0.054$), which emphasizes the value to be gained from comprehensively surveying genetic variation across all risk loci in all populations. Another possible source of the difference in association among ethnic groups could be environmental exposures that vary in frequency across populations, and which may modify the effect of these variants. However, recent studies provide little support for known breast cancer risk factors serving as modifiers of the associations with these alleles.²¹

In summary, in this multiethnic study, we evaluated the generalizability of breast cancer risk markers identified by GWAS to other populations. We observed strong evidence that, in aggregate, the 12 published risk variants are strongly associated with breast cancer risk in the majority of, but not all, populations considered. However, even for populations of non-African origin, it is clear that many more variants will be needed for this risk score to be informative in predicting breast cancer risk. Larger studies that include even more diverse populations, aimed at discovery, validation and fine-mapping, are needed to identify an accurate and more complete set of risk alleles which could better determine the contribution of these genetic regions to breast cancer risk in various populations, especially for women of African ancestry.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study is funded by grants from the National Institute of Health (CA63464, CA54281, CA098758 and CA132839) and the California Breast Cancer Research program (15UB-8402).

- Easton DF, Pooley KA, Dunning AM *et al*: Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007; **447**: 1087–1093.
- Stacey SN, Manolescu A, Sulem P *et al*: Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2007; **39**: 865–869.
- Stacey SN, Manolescu A, Sulem P *et al*: Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2008; **40**: 703–706.
- Zheng W, Long J, Gao YT *et al*: Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* 2009; **41**: 324–328.
- Thomas G, Jacobs KB, Kraft P *et al*: 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**: 579–584.
- Ahmed S, Thomas G, Ghoussaini M *et al*: Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* 2009; **41**: 585–590.
- Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA: Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* 2002; **31**: 33–36.
- Kraft P, Wacholder S, Cornelis MC *et al*: Beyond odds ratios — communicating disease risk based on genetic profiles. *Nat Rev Genet* 2009; **10**: 264–269.
- Gail MH: Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. *J Natl Cancer Inst* 2008; **100**: 1037–1041.
- Wacholder S, Hartge P, Prentice R *et al*: Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 2010; **362**: 986–993.
- Ioannidis JP: Population-wide generalizability of genome-wide discovered associations. *J Natl Cancer Inst* 2009; **101**: 1297–1299.
- Waters KM, Le Marchand L, Kolonel LN *et al*: Generalizability of associations from prostate cancer genome-wide association studies in multiple populations. *Cancer Epidemiol Biomarkers Prev* 2009; **18**: 1285–1289.
- Xu J, Kibel AS, Hu JJ *et al*: Prostate cancer risk associated loci in African Americans. *Cancer Epidemiol Biomarkers Prev* 2009; **18**: 2145–2149.
- Zheng W, Cai Q, Signorello LB *et al*: Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer Epidemiol Biomarkers Prev* 2009; **18**: 2761–2764.
- Kolonel LN, Henderson BE, Hankin JH *et al*: A Multiethnic Cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000; **151**: 346–357.
- Udler MS, Meyer KB, Pooley KA *et al*: *FGFR2* variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. *Hum Mol Genet* 2009; **18**: 1692–1703.
- Lee LG, Connell CR, Bloch W: Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res* 1993; **21**: 3761–3766.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
- Haiman CA, Hsu C, de Bakker PI *et al*: Comprehensive association testing of common genetic variation in DNA repair pathway genes in relationship with breast cancer risk in multiple populations. *Hum Mol Genet* 2008; **17**: 825–834.
- Garcia-Golas M, Hall P, Nevanlinna H *et al*: Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet* 2008; **4**: e1000054.
- Travis RC, Reeves GK, Green J *et al*: Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study. *Lancet* 2010; **375**: 2143–2151.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)