University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Starlard-Davenport, A;Allman, R;Dite, GS;Hopper, JL;Tuff, ES;Macleod, S;Kadlubar, S;Preston, M;Henry-Tillman, R

Title:
Validation of a genetic risk score for Arkansas women of color

Date:
2019-07-01

Citation:
Starlard-Davenport, A., Allman, R., Dite, G. S., Hopper, J. L., Tuff, E. S., Macleod, S., Kadlubar, S., Preston, M. & Henry-Tillman, R. (2019). Validation of a genetic risk score for Arkansas women of color. Plos One, 13 (10), https://doi.org/10.1371/journal.pone.0204834.

Persistent Link:
https://hdl.handle.net/11343/271333

License:
CC BY

PLOS ONE |

RESEARCH ARTICLE

Validation of a genetic risk score for Arkansas women of color

Athena Starlard-Davenport[1]*, Richard Allman[2], Gillian S. Dite[3], John L. Hopper[3], Erika Spaeth Tuff[4], Stewart Macleod[5], Susan Kadlubar[6], Michael Preston[7], Ronda Henry-Tillman[8]

1 Department of Genetics, Genomics & Informatics, College of Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America, **2** Genetic Technologies Ltd, Fitzroy, Victoria, Australia, **3** Centre for Epidemiology and Biostatistics, University of Melbourne, Parkville, Victoria, Australia, **4** Phenogen Sciences Inc, Charlotte, North Carolina, United States of America, **5** Genomics Core, Winthrop P. Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, Arkansas, United States of America, **6** Division of Medical Genetics, University of Arkansas for Medical Sciences, Little Rock, Arkansas, United States of America, **7** Center for Diversity Affairs and Inclusion, Winthrop P. Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, Arkansas, United States of America, **8** Department of Surgery, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, Arkansas, United States of America

* astarlar@uthsc.edu

Check for updates

OPEN ACCESS

Citation: Starlard-Davenport A, Allman R, Dite GS, Hopper JL, Spaeth Tuff E, Macleod S, et al. (2018) Validation of a genetic risk score for Arkansas women of color. PLoS ONE 13(10): e0204834. https://doi.org/10.1371/journal.pone.0204834

Editor: Amanda Ewart Toland, Ohio State University Wexner Medical Center, UNITED STATES

Received: May 7, 2018

Accepted: September 14, 2018

Published: October 3, 2018

Copyright: © 2018 Starlard-Davenport et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: A. Starlard-Davenport and R. Henry-Tillman report receiving commercial research support from Genetic Technologies Ltd. who provided support in the form of salaries for author AS-D, but did not have any additional role in the study design, data collection and analysis, or decision to publish. Phenogen Sciences Inc. did not provide any financial support, but did have an

Abstract

African American women in the state of Arkansas have high breast cancer mortality rates. Breast cancer risk assessment tools developed for African American underestimate breast cancer risk. Combining African American breast cancer associated single-nucleotide polymorphisms (SNPs) into breast cancer risk algorithms may improve individualized estimates of a woman's risk of developing breast cancer and enable improved recommendation of screening and chemoprevention for women at high risk. The goal of this study was to confirm with an independent dataset consisting of Arkansas women of color, whether a genetic risk score derived from common breast cancer susceptibility SNPs can be combined with a clinical risk estimate provided by the Breast Cancer Risk Assessment Tool (BCRAT) to produce a more accurate individualized breast cancer risk estimate. A population-based cohort of African American women representative of Arkansas consisted of 319 cases and 559 controls for this study. Five-year and lifetime risks from the BCRAT were measured and combined with a risk score based on 75 independent susceptibility SNPs in African American women. We used the odds ratio (OR) per adjusted standard deviation to evaluate the improvement in risk estimates produced by combining the polygenic risk score (PRS) with 5-year and lifetime risk scores estimated using BCRAT. For 5-year risk OR per standard deviation increased from 1.84 to 2.08 with the addition of the polygenic risk score and from 1.79 to 2.07 for the lifetime risk score. Reclassification analysis indicated that 13% of cases had their 5-year risk increased above the 1.66% guideline threshold (NRI = 0.020 (95% CI -0.040, 0.080)) and 6.3% of cases had their lifetime risk increased above the 20% guideline threshold by the addition of the polygenic risk score (NRI = 0.034 (95% CI 0.000, 0.070)). Our data confirmed that discriminatory accuracy of BCRAT is improved for African American women in Arkansas with the inclusion of specific SNP breast cancer risk alleles.

additional role in preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: We have the following interests: A. Starlard-Davenport and R. Henry-Tillman report receiving commercial research support from Genetic Technologies Ltd. R. Allman is Scientific Director at Genetic Technologies Ltd. and Erika Spaeth Tuff is employed by Phenogen Sciences Inc. There are no patents, products in development or marketed products to declare. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

Introduction

Racial disparities among women of color is evident. African American women have higher mortality rates, poorer breast cancer survival rates, and are more likely to develop early onset basal-like breast tumors that are non-responsive to hormone receptor therapy compared with European American women [1–3]. In the state of Arkansas breast cancer incidence rates have dramatically decreased due to state-funded breast and cervical cancer screening and diagnostic services for eligible Arkansas women. However, mortality rates due to breast cancer among African American Arkansas women are amongst the highest in the nation [1]. Breast cancer risk assessment tools, such as the BCRAT (also known as the Gail model) were developed by scientists at the National Cancer Institute to estimate a woman's risk of developing invasive breast cancer within the next 5 years and within her lifetime (up to age 90). The BCRAT considers race/ethnicity, age, breast cancer risk factors that influence estrogen production (such as menopausal status, body weight, parity, age at menarche, age at the time of the birth of a first child and), number of past breast biopsies, and first-degree and second-degree family history of breast cancer. Women with a 5-year risk of 1.67 percent or higher are classified as "high-risk." The BCRAT has been shown to be well calibrated for predicting breast cancer risk for European American women but unfortunately not for African American women [4, 5]. Modification of the BCRAT based on data of African American women from the Women's Contraceptive and Reproductive Experiences study also underestimated breast cancer risk in African American women [6–8].

A growing number of single-nucleotide polymorphisms (SNPs) that are associated with an increased risk of breast cancer have been identified through genome-wide association studies [9, 10]. Most of these SNPs were discovered among women of European descent. These SNPs are quite different in character from *BRCA1* and *BRCA2* mutations. While the latter confer a large increase in individual risk, they are found in very low frequencies in the population. In contrast, SNPs confer only a small increase in risk but are found in much higher frequencies in the population. While no SNP is very informative on its own, a polygenic approach to genetic screening could improve estimates of individual risk and create the possibility of individualized screening strategies [11, 12]. Several studies have shown that combining risk estimates based on SNP genotypes with risk estimates from conventional breast cancer risk prediction tools, such as the BCRAT, can improve individual risk estimates for certain ethnic populations [13–16]. Larger studies have also made efforts to identify SNPs specifically associated with breast cancer risk in African American populations through consortium including the AMBER, ROOT, and AABC [17, 18]. This approach clearly has important implications under current National Comprehensive Cancer Network guidelines for breast cancer screening and risk reduction. Such an approach has the potential to improve the efficiency of screening and chemoprevention so that more women benefit from these interventions, while at the same time reducing costs, false positives, and other harms.

The present study was initiated to confirm whether a genetic risk score based on a panel of 75 SNPs that are associated with increased breast cancer risk in women of European and/or African ancestry [19–25], could improve the BCRAT risk prediction estimates of individual breast cancer risk among a community-based cohort of African American women designed to be representative of the female population of the state of Arkansas. Our data suggests that combining a genetic risk term with a clinical risk estimate from BCRAT provides a more accurate individual risk estimate for African American women in Arkansas. These results could potentially improve targeted screening and prevention programs for African American women in Arkansas.

Materials and methods

Study population

Participants in this study were selected from a cohort of women recruited to study breast cancer in Arkansas [26]. Accrual began in 2007 and consists of 5,982 African American and 19,831 European American women aged 18 to 100 years who reside in Arkansas. Participants were recruited in several venues, including Susan G. Komen Race for the Cure events, American Cancer Society Relays for Life, community festivals and events, faith-based events, and community and business sponsored health fairs. A further 240 unaffected African American women were recruited at the Susan G. Komen Race for The Cure event held in April 2014 in Jackson, MS. Recruitment strategies at all events were the same and consisted of: (a) project introduction; (b) written informed consent; and (c) data collection. Details on study participant recruitment and establishment of the cohort was previously described [19].

Briefly, during recruitment to the cohort, baseline information on medical history, demographic factors, body weight, height, reproductive history, menstrual history, and family history of breast cancer, was obtained and a 2-mL saliva specimen from the participant was collected in an OrageneDNA Self Collection Kit. Self-reported race and ethnicity information was based on data obtained from questionnaires that were completed at recruitment. Participants also reported if a 1st degree or 2nd degree family relative (mother, sister, daughter, etc.) participated in any of the recruitment events. Participants under the age of 35 years were excluded since breast cancer risk assessment using the BCRAT algorithm is limited to women age 35 and older.

All participants in this study self-reported as African American ancestry and were selected using information in the cohort database. Study subjects included participants who had a previous diagnosis of breast cancer (n = 319), one age-matched unaffected control (n = 319) per breast cancer case, and the additional 240 unaffected controls recruited in Jackson, MS. This study and written informed consent procedure was approved by the University of Arkansas for Medical Sciences Institutional Review Board (protocol approval number 202840) and conducted in accordance with the Declaration of Helsinki.

SNP selection

The SNP panel was based upon SNPs identified as being associated with breast cancer risk by large-scale genome-wide association studies and had previously been used to derive a validated polygenic risk score using a Caucasian data set [27]. Because many of the SNPs have subsequently been found to have odds ratios of a different magnitude, or are directionally different, for African Americans, the polygenic risk score used in this study was based on the previously reported odds ratios for African Americans [18, 21]. The 75-breast cancer associated SNPs analysed in this study, together with the odds ratios calculated for the study population and the published values for both African Americans and Caucasians, are listed in S1 Table.

SNP genotyping

Genomic DNA was isolated from saliva specimens of the 319 cases and 559 controls using DNA Genotek's prepIt·L2P DNA Extraction Kit (Ontario, Canada) according to the manufacturer's instructions. Genomic DNA was evaluated and quantified using a Nanodrop UV-spectrometer (Thermo Fisher Scientific, Wilmington, DE).

Whole-genome amplification was conducted in the UAMS Genomics Core Facility using Qiagen RePLI-g Kits, which incorporates the method of multiple displacement amplification

with input of 50 nanograms of genomic DNA per reaction. Amplified samples were purified and quantified on a Nanodrop 2000 (Wilmington, DE) before being plated for genotyping.

Genotyping was performed using the Illumina GoldenGate Genotyping Assay at the Winthrop P. Rockefeller Cancer Institute's Genomics Core Facility. The GoldenGate Genotyping Assay is a flexible pre-optimized assay that uses a discriminatory DNA polymerase and ligase to interrogate 96, or from 384 to 3,072, SNP loci simultaneously. Blinded duplicate samples (5–10%) were included to assess the reproducibility of the genotypes. An average reproducibility of 99% was obtained. Individual level genotypes are presented in S2 Table.

Risk prediction

The 5-year and lifetime absolute risk of invasive breast cancer at baseline was calculated using BCRAT [6, 28]. We excluded the personal breast cancer diagnosis for cases in the risk calculations. We did not have information on biopsies or atypical hyperplasia, so these variables were coded as 'unknown', which is an accepted standard response option for BCRAT. In accordance with BCRAT's design, its application was restricted to women aged 35 years or older [6, 28]. We incorporated the modified Women's Contraceptive and Reproductive Experiences (WeCARE) model for breast cancer risk estimation in African American women.

SNP-based risk score and combined risk score

We used the approach of Mealiffe et al. [13] to calculate a SNP-based (relative) risk score. We determined the SNP-based risk score using estimates of the odds ratios (OR) per allele and the risk allele frequencies (p) from a previously published study [27], assuming independent and additive risks on the log OR scale. For each SNP, we calculated the unscaled population average risk as $\mu = (1 - p)^2 + 2p\,(1 - p)\,\mathrm{OR} + p^2\mathrm{OR}^2$ as previously described [25]. After the adjusted risk values were calculated [25], we calculated the overall SNP-based risk score by multiplying the adjusted risk values for each of the 75 SNPs investigated [13]. Lastly, the combined risk score was calculated by multiplying the SNP-based score by the model's predicted 5-year/lifetime risk of breast cancer as described by Dite et al [25].

Statistical analysis

We log transformed the BCRAT risk score, SNP-based score, and combined risk score for all analyses. To test for associations between the BCRAT risk score, the SNP-based risk score, and the combined risk score, we used Pearson correlation.

We used logistic regression to estimate risk associations for the BCRAT, SNP-based, and combined risk scores in terms of the OR per adjusted standard deviation (OPERA), where the log 5-year predicted risk was adjusted for age [29]. We assessed model calibration using the Hosmer–Lemeshow goodness-of-fit test And measured discrimination between cases and controls was measured using the area under the receiver operating characteristic (AUC) curves of the risk scores as previously described [25]. We categorized 5-year absolute risks as low risk ($<$1.66%) and high risk ($\geq$1.66%) and constructed reclassification tables as a cross-tabulation of the classification of the BCRAT risk score with the classification of the combined risk score. Similarly, we categorized lifetime absolute risks as low risk ($<$20%) and high risk ($\geq$20%) and constructed reclassification tables as a cross-tabulation of the classification of the BCRAT risk score with the classification of the combined risk score from. We calculated the net reclassification improvement statistic as $P(\text{up}|\text{case}) - P(\text{down}|\text{case}) + P(\text{down}|\text{control}) - P(\text{up}|\text{control})$ as previously described [25]. We used Stata Release 13 [30] for two sided statistical tests and all statistical analyses. A P value less than 0.05 was considered statistically significant.

Results

Table 1 shows the characteristics of the study participants and the questionnaire variables used in the calculation of the BCRAT risk score. For cases, the mean age at diagnosis was 54.6 years

Table 1. Characteristics of study participants.

	Cases (n = 319)		Controls (n = 559)	
	N	(%)	N	(%)
Age (years)				
35–39	23	(7.2)	39	(7)
40–44	34	(10.7)	69	(12.3)
45–49	48	(15.1)	88	(15.7)
50–54	61	(19.1)	127	(22.7)
55–59	49	(15.4)	101	(18.1)
60–64	48	(15.1)	53	(9.5)
65–69	29	(9.1)	46	(8.2)
70–74	15	(4.7)	21	(3.8)
≥70	12	(3.8)	15	(2.7)
Number of first-degree relatives with breast cancer				
0	177	(55.5)	336	(60.1)
1	45	(14.1)	55	(9.8)
≥2	49	(15.4)	5	(0.9)
Missing	48	(15.1)	163	(29.2)
Age at menarche (years)				
≤9	6	(1.9)	25	(4.5)
10	30	(9.4)	53	(9.5)
11	34	(10.7)	55	(9.8)
12	94	(29.5)	166	(29.7)
13	84	(26.3)	118	(21.1)
14	23	(7.2)	55	(9.8)
15	12	(3.8)	32	(5.7)
16	16	(5)	26	(4.7)
≥17	7	(2.2)	15	(2.7)
Missing	13	(4.1)	14	(2.5)
Age at first live birth (years)				
≤19	106	(33.2)	181	(32.4)
20–24	105	(32.9)	182	(32.6)
25–29	40	(12.5)	66	(11.8)
30–34	14	(4.4)	37	(6.6)
≥35	6	(1.9)	8	(1.4)
Missing	48	(15.1)	85	(15.2)
Number of breast biopsies				
0	0	0	31	(5.6)
1	12	(3.8)	7	(1.3)
2	5	(1.6)	2	(0.4)
Missing	302	(94.7)	519	(92.8)
Had hyperplasia				
No	7	(2.2)	37	(6.6)
Yes	12	(3.8)	4	(0.7)
Missing	300	(94)	518	(92.7)

https://doi.org/10.1371/journal.pone.0204834.t001

Table 2. AUC for the unadjusted risk scores.

Log-transformed risk score	AUC	(95% CI)
5-year BCRAT	0.678	(0.639, 0.716)
Lifetime BCRAT	0.652	(0.612, 0.692)
SNP	0.581	(0.541, 0.620)
5-year BCRAT × SNP	0.679	(0.642, 0.716)
Lifetime BCRAT × SNP	0.658	(0.619, 0.696)

Change in AUC (1 degree of freedom).

5-year BCRAT: $\chi^2 = 0.0$, $P = 1.0$.

Lifetime BCRAT: $\chi^2 = 0.13$, $P = 0.7$.

$P < 0.05$ is considered statistically significant.

https://doi.org/10.1371/journal.pone.0204834.t002

(standard deviation [SD] = 10.4), and for controls, the mean age at interview was 53.3 years (SD = 9.8). The mean BCRAT 5-year risk of breast cancer was 1.80% (SD 1.10%) for cases and 1.20% (SD 0.50%) for controls. The mean BCRAT lifetime risk of breast cancer was 10.4% (SD 5.20%) for cases and 8.0% (SD 2.70%) for controls. The mean SNP-based risk score was 1.02 (SD 0.43) for cases and 0.89 (SD 0.38) for controls. S1 Table shows, for each of the 75 SNPS, the per-allele OR and 95% CI in this study as well as the previously published per-allele OR and 95% CI.

The cases in the study population were oversampled for family history; in particular 49 (15.4%) of cases had 2 or more first-degree relatives, well in excess of that observed in population-based case samples [31], This resulted in an AUC value for BCRAT which was substantially higher than previously published values, a consequence of number of affected relatives being included in the BCRAT algorithm. For the 5-year risk estimates, the combined risk score did not improve the AUC over BCRAT alone (p = 0.4), similarly for lifetime risk, the combined score did not improve the AUC (p = 0.7; see Table 2 and Fig 1).

The OPERA estimates of the OR, however, did confirm that the combined risk scores for 5-year and lifetime risk gave greater discrimination than the SNP-based risk score and the corresponding BCRAT 5-year and lifetime risk scores alone (Table 3). Furthermore, incorporation of the SNP-based risk score improved the OPERA estimate from 1.84 to 2.08 for the BCRAT 5-year risk estimate and from 1.79 to 2.07 for the BCRAT lifetime risk estimate.

Reclassification tables were used to calculate the number of cases and controls whose risk classification changed with respect to the American Society of Clinical Oncology's 5-year risk threshold of 1.66% for Tamoxifen chemoprevention and the American Cancer Society's lifetime risk threshold of 20% for MRI screening is recommended (Tables 4 and 5). Reclassification analysis showed that after incorporating the SNP-based risk score, 13% of cases had their 5-year risk increased above the 1.66% threshold (NRI = 0.020 (95% CI -0.040, 0.080)) and 6.3% of cases had their lifetime risk increased above the 20% threshold (NRI = 0.034 (95% CI 0.000, 0.070)).

Discussion

Genome-wide association studies conducted mainly in European or Asian populations have identified an increasing number of genetic susceptibility loci for breast cancer. In addition to the observation that SNPs are associated with increased risk of developing breast cancer, several studies have investigated the possibility of combining the SNP-based risk scores with conventional breast cancer risk prediction tools, particularly for women of European descent [13–16, 32] but also recently for women of African ancestry and Hispanic ancestry [33]. Using the

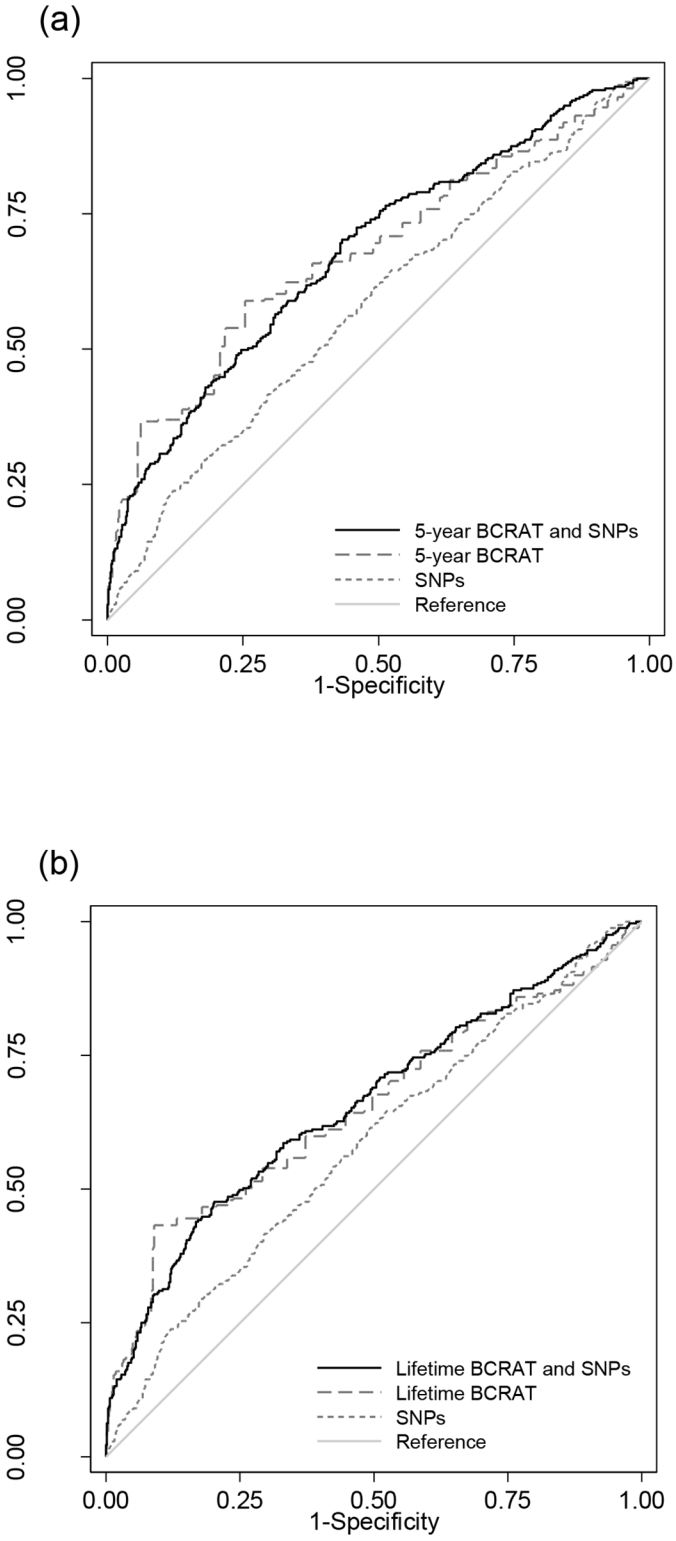

Fig 1. The receiver operating characteristic curves (ROC) and area under the curve (AUC) for a) 5-year BCRAT risk, single-nucleotide polymorphism (SNP) risk, and combined 5-year BCRAT × SNP risk and b) lifetime risk.

https://doi.org/10.1371/journal.pone.0204834.g001

Table 3. Age-adjusted OPERA estimates of the OR, with corresponding 95% Cis.

Log-transformed risk score	OPERA OR	(95% CI)	P
5-year BCRAT	1.835	(1.611, 2.090)	< 0.001
Lifetime BCRAT	1.798	(1.589, 2.035)	< 0.001
SNP	1.434	(1.220, 1.687)	< 0.001
5-year BCRAT × SNP	2.085	(1.774, 2.449)	< 0.001
Lifetime BCRAT × SNP	2.069	(1.757, 2.438)	< 0.001

$P < 0.05$ is considered statistically significant.

https://doi.org/10.1371/journal.pone.0204834.t003

Table 4. Five-year BCRAT reclassification table.

5-year BCRAT	Combined risk score	
	<1.66%	≥1.66%
Cases		
<1.66%	147	42
≥1.66%	34	96
Controls		
<1.66%	411	47
≥1.66%	44	57

NRI = 0.020 (95% CI -0.040, 0.080).

Abbreviation: NRI, net reclassification improvement.

https://doi.org/10.1371/journal.pone.0204834.t004

same SNPs validated in a recent comprehensive panel for African Americans [25], this study's smaller, independent cohort of African American women in Arkansas confirms the added value of SNP to breast cancer risk assessment among the African American population. Importantly, for European ancestry women, the combined SNP-based and risk prediction tool scores are now among the strongest known measures for differentiating women with and without breast cancer [32]. Given the disparity in stage at breast cancer diagnosis between African American women and other US populations, identification and testing the validity of a breast cancer risk assessment tool that may increase screening and prevention awareness in African American women is needed and SNP-based risk assessment may enable the medical

Table 5. Lifetime BCRAT reclassification table.

Lifetime BCRAT	Combined risk score	
	<20%	≥20%
Cases		
<20%	279	20
≥20%	8	12
Controls		
<20%	554	3
≥20%	1	1

NRI = 0.034 (95% CI 0.000, 0.070).

Abbreviation: NRI, net reclassification improvement.

https://doi.org/10.1371/journal.pone.0204834.t005

community to better identify those women that would benefit from increased screening, screening adherence or risk reducing medication.

We have evaluated the effect of combining a polygenic risk score with the BCRAT risk score (incorporating the modified Women's Contraceptive and Reproductive Experiences WeCARE model) for breast cancer risk estimation in African American women. SNP-based risk has been shown to be largely independent of family history with only a modest attenuation when adjusting for family history [27]. In this study, the AUC did not significantly improve for the 5-year or lifetime risk estimates by the addition of the SNP-based risk score. Receiver operating characteristic curve analysis has previously been criticized in the context of risk prediction test performance [13, 34], and it is our view that the AUC values obtained for the BCRAT in this study were inflated due to the enriched and strong family history of the cases resulting in an AUC substantially higher than that previously reported for African American women unselected for family history [33].

We believe a better way to combine risk factors when they are measured using different scales is to utilize methodologies in which the standard deviation of the risk gradient is incorporated into the analysis. We have therefore, preferentially utilized the odds per adjusted standard deviation method (OPERA), as described by Hopper 2016 [29]. The concept of OPERA allows comparison of the ability of risk factors to differentiate cases from controls on a population, as distinct from individual, basis and can be applied to continuous, binary, and ordinal risk factors. Following the OPERA concept, we have presented the OR in terms of the age-adjusted standard deviation of the risk score. OPERA is based on the fact that the estimated risk gradient for a risk factor is its change when holding constant all other factors that have been controlled for by design or analysis. That is why we fitted the age-adjusted risk scores. If the OR is the risk gradient on a given scale, then OPERA = OR^s, where s is the estimated SD of the risk factor in the population after adjusting for all other factors (and can be estimated using the controls). Under these terms of reference, the addition of the SNP-based risk score resulted in a 13% improvement in the OPERA for the BCRAT 5-year risk estimate and a 15% improvement for the BCRAT lifetime risk estimate. This improvement in risk assessment might help the medical community begin to tackle the differences in stage at diagnosis; significantly fewer Non-Hispanic blacks are diagnosed at a localized stage compared to non-Hispanic whites [35].

We acknowledge that a limitation of the present study included small participant sample size and lack of data pathological subtypes of breast cancer among the participants. Within the cohort of women used for this study population, many were also unable to complete the BCRAT in its entirety (Table 1). However, because the outcome of these data focuses on the change as a result of the incorporation of SNP-based risk, the baseline missing BCRAT data would not significantly alter the combined effects of SNP-based change. However, as identified above, the cases are enriched for family history, which has impacted the BCRAT test performance.

Reclassification of individual risk is important if the reclassification results in an individual's risk exceeding clinical guideline thresholds, such as those for chemoprevention or screening. Reclassification analysis showed that 22% of cases that were previously below the American Society of Clinical Oncology's risk threshold of 1.66% for Tamoxifen chemoprevention would have their individual risk elevated above the threshold after incorporation of the SNP-based risk score. Similarly, 6.7% of cases who were previously below the 20% lifetime risk threshold for MRI screening would have their risk elevated above the threshold by the addition of the SNP-based risk score.

The unique feature of our population-based cohort is that the women are representative of 75 counties, encompassing both rural and urban areas, within the state of Arkansas [19].

Arkansas has one of the highest breast cancer mortality rates among women of color [1] and; therefore, our results could potentially improve targeted screening and prevention programs for African American women in Arkansas.

Conclusions

In conclusion, our results indicate that discriminatory accuracy of BCRAT is improved for African American women in Arkansas with the inclusion of African American specific SNP breast cancer risk alleles. We anticipate that the SNP-based risk score will further improve when genome-wide association studies use Phase I data sets that are relevant to the African American population, and when fine mapping studies have been conducted within that population.

Supporting information

S1 Table. Unadjusted per-allele ORs for individual SNPs and the previously published ORs used to derive the risk scores.
(DOCX)

S2 Table. Individual level genotypes of participants in our study.
(CSV)

Acknowledgments

The authors thank Dr. Phillip Ley of Madison MS for facilitating the collection of samples at the Susan G Komen "Race for the Cure" Event in April 2014.

Author Contributions

Conceptualization: Athena Starlard-Davenport, Richard Allman, Ronda Henry-Tillman.

Data curation: Gillian S. Dite, John L. Hopper, Stewart Macleod, Susan Kadlubar.

Formal analysis: Athena Starlard-Davenport, Gillian S. Dite, John L. Hopper, Stewart Macleod.

Funding acquisition: Richard Allman.

Investigation: Athena Starlard-Davenport, Stewart Macleod, Susan Kadlubar.

Methodology: Gillian S. Dite, John L. Hopper.

Project administration: Athena Starlard-Davenport, Susan Kadlubar, Michael Preston, Ronda Henry-Tillman.

Resources: John L. Hopper, Stewart Macleod, Susan Kadlubar, Ronda Henry-Tillman.

Software: John L. Hopper.

Supervision: Athena Starlard-Davenport, Ronda Henry-Tillman.

Validation: Athena Starlard-Davenport, Gillian S. Dite, John L. Hopper, Stewart Macleod.

Visualization: Athena Starlard-Davenport, Gillian S. Dite, John L. Hopper.

Writing – original draft: Athena Starlard-Davenport, Richard Allman, Ronda Henry-Tillman.

Writing – review & editing: Athena Starlard-Davenport, Richard Allman, John L. Hopper, Erika Spaeth Tuff, Ronda Henry-Tillman.

References

1. DeSantis C, Ma J, Bryan L, and Jemal A. Breast cancer statistics, 2013. *CA Cancer J Clin*. (2014); 64: 52–62 https://doi.org/10.3322/caac.21203 PMID: 24114568
2. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. (2006); 295: 2492–502 https://doi.org/10.1001/jama.295.21.2492 PMID: 16757721
3. Perez CA, Zumsteg ZS, Gupta G, Morrow M, Arnold B, Patil SM, et al. Black race as a prognostic factor in triple-negative breast cancer patients treated with breast-conserving therapy: a large, single-institution retrospective analysis. *Breast Cancer Res Treat*. (2013); 139: 497–506 https://doi.org/10.1007/s10549-013-2550-x PMID: 23645006
4. Costantino JP, Gail MH, Pee D, Anderson S, Redmond CK, Benichou J, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst*. (1999); 91: 1541–8 PMID: 10491430
5. Rockhill B, Spiegelman D, Byrne C, Hunter DJ, and Colditz GA. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst*. (2001); 93: 358–66 PMID: 11238697
6. Gail MH, Costantino JP, Pee D, Bondy M, Newman L, Selvan M, et al. Projecting individualized absolute invasive breast cancer risk in African American women. *J Natl Cancer Inst*. (2007); 99: 1782–92 https://doi.org/10.1093/jnci/djm223 PMID: 18042936
7. Adams-Campbell LL, Makambi KH, Palmer JR, and Rosenberg L. Diagnostic accuracy of the Gail model in the Black Women's Health Study. *Breast J*. (2007); 13: 332–6 https://doi.org/10.1111/j.1524-4741.2007.00439.x PMID: 17593036
8. Adams-Campbell LL, Makambi KH, Frederick WA, Gaskins M, Dewitty RL, and McCaskill-Stevens W. Breast cancer risk assessments comparing Gail and CARE models in African-American women. *Breast J*. (2009); 15 Suppl 1: S72–5.
9. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. (2007); 447: 1087–93. https://doi.org/10.1038/nature05887 PMID: 17529967
10. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*. (2013); 45: 353–61. https://doi.org/10.1038/ng.2563 PMID: 23535729
11. Pharoah PD, Antoniou AC, Easton DF, and Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med*. (2008); 358: 2796–803 https://doi.org/10.1056/NEJMsa0708739 PMID: 18579814
12. Maas P, Barrdahl M, Joshi AD, Auer PL, Gaudet MM, Milne RL, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA Oncol*. (2016);
13. Mealiffe ME, Stokowski RP, Rhees BK, Prentice RL, Pettinger M, and Hinds DA. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst*. (2010); 102: 1618–27. https://doi.org/10.1093/jnci/djq388 PMID: 20956782
14. Comen E, Balistreri L, Gonen M, Dutra-Clarke A, Fazio M, Vijai J, et al. Discriminatory accuracy and potential clinical utility of genomic profiling for breast cancer risk in BRCA-negative women. *Breast Cancer Res Treat*. (2011); 127: 479–87. https://doi.org/10.1007/s10549-010-1215-2 PMID: 20957429
15. Dite GS, Mahmoodi M, Bickerstaffe A, Hammet F, Macinnis RJ, Tsimiklis H, et al. Using SNP genotypes to improve the discrimination of a simple breast cancer risk prediction model. *Breast Cancer Res Treat*. (2013); 139: 887–96. https://doi.org/10.1007/s10549-013-2610-2 PMID: 23774992
16. Brentnall AR, Evans DG, and Cuzick J. Value of phenotypic and single-nucleotide polymorphism panel markers in predicting the risk of breast cancer. *J Genet Syndr Gene Ther*. (2013); 4: 202
17. Murphy ME, Liu S, Yao S, Huo D, Liu Q, Dolfi SC, et al. A functionally significant SNP in TP53 and breast cancer risk in African-American women. *NPJ Breast Cancer*. (2017); 3: 5. https://doi.org/10.1038/s41523-017-0007-9 PMID: 28649645
18. Feng Y, Rhie SK, Huo D, Ruiz-Narvaez EA, Haddad SA, Ambrosone CB, et al. Characterizing Genetic Susceptibility to Breast Cancer in Women of African Ancestry. *Cancer Epidemiol Biomarkers Prev*. (2017); 26: 1016–1026. https://doi.org/10.1158/1055-9965.EPI-16-0567 PMID: 28377418

19. Siddiq A, Couch FJ, Chen GK, Lindstrom S, Eccles D, Millikan RC, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum Mol Genet*. (2012); 21: 5373–84. https://doi.org/10.1093/hmg/dds381 PMID: 22976474

20. Feng Y, Stram DO, Rhie SK, Millikan RC, Ambrosone CB, John EM, et al. A comprehensive examination of breast cancer risk loci in African American women. *Hum Mol Genet*. (2014); 23: 5518–26. https://doi.org/10.1093/hmg/ddu252 PMID: 24852375

21. Palmer JR, Ruiz-Narvaez EA, Rotimi CN, Cupples LA, Cozier YC, Adams-Campbell LL, et al. Genetic susceptibility loci for subtypes of breast cancer in an African American population. *Cancer Epidemiol Biomarkers Prev*. (2013); 22: 127–34. https://doi.org/10.1158/1055-9965.EPI-12-0769 PMID: 23136140

22. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. *J Natl Cancer Inst*. (2015); 107.

23. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. (2007); 447: 1087–93. https://doi.org/10.1038/nature05887 PMID: 17529967

24. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet*. (2008); 40: 703–6 https://doi.org/10.1038/ng.131 PMID: 18438407

25. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet*. (2010); 42: 504–7. https://doi.org/10.1038/ng.586 PMID: 20453838

26. Bondurant KL, Harvey S, Klimberg S, Kadlubar S, and Phillips MM. Establishment of a southern breast cancer cohort. *Breast J*. (2011); 17: 281–8 https://doi.org/10.1111/j.1524-4741.2011.01082.x PMID: 21489034

27. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants [published online ahead of print April 8, 2015]. *J Natl Cancer Inst*. (2015); 107: https://doi.org/10.1093/jnci/djv036 PMID: 25855707

28. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. (1989); 81: 1879–86 PMID: 2593165

29. Hopper JL. Odds PER Adjusted standard deviation (OPERA): Comparing strengths of associations for risk factors measured on different scales, and across diseases and populations. *Am J Epidemiol*. (2015); 182: 863–867

30. StataCorp, *Stata statistical software*, *release 13*. 2013, StataCorp LP: College Station, TX.

31. Dite GS, Jenkins MA, Southey MC, Hocking JS, Giles GG, McCredie MR, et al. Familial risks, early-onset breast cancer, and BRCA1 and BRCA2 germline mutations. *J Natl Cancer Inst*. (2003); 95: 448–57 PMID: 12644538

32. Dite GS, MacInnis RJ, Bickerstaffe A, Dowty JG, Allman R, Apicella C, et al. Breast cancer risk prediction using clinical models and 77 independent risk-associated SNPs for women aged under 50 years: Australian Breast Cancer Family Registry. *Cancer Epidemiology*, *Biomarkers & Prevention*. (2015); https://doi.org/10.1158/1055-9965.EPI-15-0838 PMID: 26677205

33. Allman R, Dite GS, Hopper JL, Gordon O, Starlard-Davenport A, Chlebowski R, et al. SNPs and breast cancer risk prediction for African American and Hispanic women. *Breast Cancer Res Treat*. (2015); 154: 583–589 https://doi.org/10.1007/s10549-015-3641-7 PMID: 26589314

34. Cook NR. Use and misuse of the receiver operating characterisitic curve in risk prediction. *Circulation* (2007); 115: 191–195

35. Society AC. Cancer Facts & Figures for frican Americans 2016–1018. Atlanta: American Cancer Society. (2016);