BBA Clinical 2 (2014) 94–102

Contents lists available at ScienceDirect

BBA Clinical

journal homepage: http://www.journals.elsevier.com/bba-clinical/

A steroid metabolizing gene variant in a polyfactorial model improves risk prediction in a high incidence breast cancer population

Eldon R. Jupe [a,1], Kathie M. Dalessandri [b], John J. Mulvihill [c], Rei Miike [d], Nicholas S. Knowlton [e], Thomas W. Pugh [a,2], Lue Ping Zhao [f], Daniele C. DeFreese [a,3], Sharmila Manjeshwar [a,4], Bobby A. Gramling [a,5], John K. Wiencke [d], Christopher C. Benz [g,h,*]

[a] Research and Development, InterGenetics Incorporated, Oklahoma City, OK, USA
[b] Surgeon-Scientist, Point Reyes Station, CA, USA
[c] Department of Pediatrics, Section of Genetics, University of Oklahoma, Oklahoma City, OK, USA
[d] Department of Neurological Surgery, University of California, San Francisco, CA, USA
[e] NSK Statistical Solutions LLC, Choctaw, OK, USA
[f] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
[g] Division of Hematology-Oncology, University of California, San Francisco, CA, USA
[h] Buck Institute for Research on Aging, Novato, CA, USA

ARTICLE INFO

Article history:
Received 26 September 2014
Received in revised form 29 October 2014
Accepted 2 November 2014
Available online 8 November 2014

Keywords:
Breast cancer
Polyfactorial risk model (PFRM)
Single nucleotide polymorphisms (SNPs)
Aldosterone synthase variant (*CYP11B2*, -344T/C)

ABSTRACT

Background: We have combined functional gene polymorphisms with clinical factors to improve prediction and understanding of sporadic breast cancer risk, particularly within a high incidence Caucasian population.
Methods: A polyfactorial risk model (PFRM) was built from both clinical data and functional single nucleotide polymorphism (SNP) gene candidates using multivariate logistic regression analysis on data from 5022 US Caucasian females (1671 breast cancer cases, 3351 controls), validated in an independent set of 1193 women (400 cases, 793 controls), and reassessed in a unique high incidence breast cancer population (165 cases, 173 controls) from Marin County, CA.
Results: The optimized PFRM consisted of 22 SNPs (19 genes, 6 regulating steroid metabolism) and 5 clinical risk factors, and its 5-year and lifetime risk prediction performance proved significantly superior (~2-fold) over the Gail model (Breast Cancer Risk Assessment Tool, BCRAT), whether assessed by odds (OR) or positive likelihood (PLR) ratios over increasing model risk levels. Improved performance of the PFRM in high risk Marin women was due in part to genotype enrichment by a *CYP11B2* (-344T/C) variant.
Conclusions and general significance: Since the optimized PFRM consistently outperformed BCRAT in all Caucasian study populations, it represents an improved personalized risk assessment tool. The finding of higher Marin County risk linked to a *CYP11B2* aldosterone synthase SNP associated with essential hypertension offers a new genetic clue to sporadic breast cancer predisposition.

1. Introduction

Breast cancer continues to be a common cancer in US women with a lifetime risk of ~12% (1 in 8), with an ever-increasing and overall (age adjusted) incidence of ~127 per 100,000 for non-Hispanic Caucasian women [1]. The annual breast cancer-specific mortality rate, declining slightly due to early detection and treatment advances, is still ~25 per 100,000 overall, with a 5-year breast cancer-specific death rate of 14% for non-Hispanic Caucasian women, and remains the leading cause of cancer deaths for all women age 40–55. A key to improving breast cancer survival is early detection [2], achieved in part by identifying new risk factors and better models for estimating individual breast cancer risk [3,4]. Improvements in individualized risk estimation would allow more accurate identification of those women most likely to benefit from regular screening with more sensitive methods or from more aggressive prevention strategies [5,6].

The National Cancer Institute Breast Cancer Risk Assessment Tool (BCRAT), or Gail model, and its updated versions are the most commonly used tools for breast cancer risk estimation [7,8], with attention turning to methods that might improve upon its predictive accuracy. The

* Corresponding author at: Buck Institute for Research on Aging, 8001 Redwood Blvd., Novato, CA 94945, USA. Tel.: +1 415 209 2092.
E-mail address: cbenz@buckinstitute.org (C.C. Benz).
[1] Current address: Analytical Edge Discoveries, Oklahoma City, OK, USA.
[2] Current address: Consultant Analytical Edge Discoveries, Oklahoma City, OK, USA.
[3] Current address: Analytical Edge Discoveries, Oklahoma City, OK, USA.
[4] Current address: Genoptix Medical Laboratory, Carlsbad, CA, USA.
[5] Current address: Integris Mayes County Medical Center, Pryor, OK, USA.

http://dx.doi.org/10.1016/j.bbacli.2014.11.001

BCRAT has been demonstrated to be well-calibrated in its ability to estimate the number of cancers likely to emerge in a population of women seeking regular mammography screening [9–11], but, for individual patient counseling, it lacks the desired discriminatory accuracy [12,13]. Incremental improvements in the BCRAT have been achieved by the addition of additional clinical risk factors including fine needle aspirate cytology [14] as well as mammographic density and weight [15]; however, such modifications have not been adopted for widespread use. Common variants in candidate genes with probable physiological roles in pathways involved in breast carcinogenesis have long been studied for association with breast cancer [16,17]. Independently, genome wide association studies (GWASs) have identified a core group of 7–10 single nucleotide polymorphisms (SNPs) associated with breast cancer, most of which do not have any known functional consequence or obvious role in the disease process [18–22]. For both candidate and GWAS identified SNPs the risks conferred by any one gene variant are small to modest (OR = 1.2–1.5) and cannot be used to effectively determine risk. Investigators have suggested that breast cancer risk for the majority of women (including familial associated but BRCA1/2-negative cancers) is likely to be polygenic [23–26], and several studies have utilized estimates of relative risks and allele frequencies for the GWAS SNPs to determine if multiplicative risk estimates for the SNPs alone or SNPs multiplied by BCRAT risks improve risk estimation [27–31]. These studies have reported minimal improvements in risk prediction.

In order to develop a new polyfactorial breast cancer risk assessment model (PFRM), we have taken a candidate gene approach [26,32], similar to that used for the successful development of multigene assays routinely used to predict the likelihood of developing a distant recurrence in a newly diagnosed breast cancer patient [34], and built an age-specific model that integrates these genetic factors with known clinical breast cancer risk factors [33]. Common functional SNPs in candidate genes known or likely to influence breast cancer development were first identified from the published literature and genomic databases. Genotyping and clinical risk factor data were then combined for each individual in large model building and validation case–control datasets consisting of participants enrolled from multiple geographic regions within the US. A multi-step statistical protocol employing multivariate logistic regression was used to build the final model, and the performance of this optimized PFRM was evaluated relative to the widely used BCRAT, and then reassessed using another independent dataset of DNA samples and clinical risk data from case–control participants enrolled from a very high incidence breast cancer population in Marin County, CA [35,36].

2. Materials and methods

2.1. Model building and validation study populations

As described previously, the study population consisted of women in a case–control study conducted from 1996 through 2006, in Oklahoma City (OK), Seattle (WA), San Diego (CA), Kansas City (KS and MO), Orlando (FL), and Charleston (SC) [26,32]. At each site, most women approached were enrolled in the study; cases self-reported a diagnosis of breast cancer at any time, and controls reported no diagnosis of any cancer. Participants at mammography clinics were either newly diagnosed (or follow-up) cases, or were cancer-free controls based on history and screening. Cases were also enrolled in surgery and oncology clinics with controls obtained in general practice clinics in the same or a nearby medical facility. Participants also enrolled at community-based events, such as Komen Races for the Cure. At each site both cases and controls were enrolled. All participants gave informed consent and completed a questionnaire providing information on approximately 50 risk factors including their medical history, family history of cancer, and lifestyle factors; in addition they provided a buccal cell sample in commercial mouthwash. The initial model building study set consisted of 5022 Caucasian females: 1671 breast cancer cases and 3351 age-matched cancer-free controls, and the validation study set consisted of 1193 Caucasian females: 400 breast cancer cases and 793 age-matched cancer-free controls.

2.2. High incidence Marin study population

This population-based case–control study has previously been described [35,36]. Briefly, eligible cases included any female resident of Marin County diagnosed with primary breast cancer between 1997 and 1999. The 285 cases identified with breast cancer were matched by age at diagnosis and ethnicity (non-Hispanic Caucasian) to 286 eligible controls selected by random digit dialing. All enrolled and consented participants completed a comprehensive questionnaire about lifestyle, reproductive and clinical risk factors (including personal medical history and family history of breast cancer). Of note, a prior report of this study based on the questionnaire data showed no significant differences between the case and control study groups for common breast cancer risk factors such as those used in BCRAT, including age, age at menarche, age at first live birth, number of first-degree relatives with breast cancer, history and outcome of previous breast biopsies [36]. With the exception of alcohol consumption, additional risk factors including use of hormone replacement therapy, prior therapeutic exposure to radiation, as well as residence time in Marin County were found to be statistically unassociated with breast cancer risk [36]. The majority of participants completing the Marin questionnaire, including 164 cases and 174 controls, also donated buccal samples which were initially cryobanked for later DNA analysis. The Investigational Review Board at University of California, San Francisco approved the informed consent process, design and protocol for this Marin study.

2.3. DNA isolation and genotyping

Processing of all coded buccal samples for DNA extraction and genotyping was performed blinded to case or control study assignment. Genomic DNA was isolated by a Gentra PureGene DNA purification kit (Gentra, Minneapolis, MN), and genotyping was performed using microbead-based allele-specific primer extension (ASPE) followed by analysis on the Luminex 100™ (Luminex, Inc. Austin, TX) as previously described [26,32]. For all assays, at least 5% of the specimens were genotyped in duplicate with a concordance rate of >99%.

2.4. Selection of candidate gene polymorphisms (SNPs)

Coded DNA samples from the model building and validation study populations were genotyped for 117 common, functional polymorphisms from 87 distinct candidate genes (Supplementary Table 1). These SNPs were selected from thousands of candidates considered from published papers, reviews and meta-analyses as well as genomic cancer databases such as Online Mendelian Inheritance in Man (OMIM) and the Cancer Genome Anatomy Project (CGAP). The candidate gene markers used for model building were ultimately narrowed by applying the following selection criteria: (1) associated with risk of breast or other cancers in at least one peer-reviewed publication, or having a plausible physiological role in a major pathway implicated in breast carcinogenesis; (2) demonstrated or predicted to have functional physiological consequences, including non-synonymous amino acid substitutions in protein-coding regions leading to alterations in enzymatic activity or regulatory sequence changes (promoter or 3′UTR); and (3) a minor allele common in major ethnic groups. For Caucasians, the minor allele frequencies ranged between 0.01 and 0.50, with a median and mean of 0.30 and 0.28, respectively. These SNP candidates were in genes encoding hormone receptors, extra-cellular matrix proteins, immune modulators, modulators of oxidative potential, growth factors and signaling molecules, as well as proteins involved in synthesis and metabolism of steroid hormones and related molecules, DNA repair and metabolism, cell cycle control and apoptosis.