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.

2.5. Model building strategy

Demographic variables and allelic frequencies were very similar in different geographic areas (Supplemental Table 1) [32]. Therefore, all sample sources were combined to derive the model building and validation study sets. The combined model building and validation datasets contained 2071 Caucasian cases and 4144 controls between the ages of 30 and 69, with age for cases being their age at diagnosis and age for controls being their age at the time of study enrollment. The combined dataset was randomly divided, with 80% for model building and the remaining 20% for validation. Clinical risk factors for each set of participants were similar (Supplementary Table 2); both model building and validation DNA samples were kept entirely separate. The genotype frequencies in the general population at steady state are expected to be in Hardy Weinberg Equilibrium (HWE). Before its use in model building, the goodness-of-fit χ^2-test was used to confirm HWE [37]. Ten of the 117 candidate SNPs did not conform to HWE expectations ($p \leq 0.05$) and were excluded from further analyses. The primary analytical goal was to systematically evaluate genotypic and clinical/lifestyle factor associations with case–control status using multivariate logistic regression (MLR) [38]. Our model building strategy took into account the fact that breast cancer is a complex disease where the effect of a specific SNP could vary with age [32,33]. Thus, the modeling analyses evaluated terms in both age-invariant and age-interactive manners for their contribution to risk estimation, and considered effects of SNPs alone, SNP–age interactions, and SNP–SNP interactions. To focus model building on the SNPs most likely to contribute in a multivariate setting while simultaneously reducing the dimensionality of the problem [39], we employed a feature selection strategy to reduce the number of eligible terms [40]. First of all, we chose the top 25% of SNPs [41,42] based on a univariate χ^2 p-value. Following feature selection, the reduced dataset was modeled with a forward stepwise selection method with the selection p-value of 0.1 and the exit p-value of 0.05. To avoid reliance on asymptotic theories for inference, we utilized a bootstrap method for computing standard errors. Each step in the process was bootstrapped 5000 times [43]. The maximum number of steps allowed was 100. This analysis was initially performed on the entire model building dataset to identify informative terms for ages 30 through 69. Published analyses of several candidate SNPs had demonstrated both “pre-” and “post-” menopause specific associations when stratified at age 50 or by first-degree relative status [32,44–51]. To capture these complexities, age strata (30–49 and 49–69) were also analyzed stratified on presence or absence of at least one first-degree relative with breast cancer. To keep our models parsimonious, informative terms identified for ages 30–69 were not included as candidate terms in subsequent analyses. Additional informative terms were identified for the 30–49 age stratum both with and without family history of breast cancer. However, analyses of the age 50–69 group did not identify additional informative terms. The informative terms identified overall (30–69) and within each strata of 30–49 year olds (by family history) were combined and maximum likelihood estimates were used to produce a single integrated PFRM which was used to compute an individual relative risk. Using SEER breast cancer incidence and competing mortality rates [52] by the described methods [7,53], the PFRM was used to estimate the probability of developing breast cancer from the individuals' current age over the next 5 years and to estimate lifetime risk up to age 90.

2.6. Performance comparison and calibration of optimized PFRM relative to BCRAT

The performance of the optimized PFRM was examined in comparison to the BCRAT alone. Model performance was evaluated for women ≥35 years of age in the model building and validation populations. The OR was used to compare the proportion of cases and controls at various risk thresholds output by each model (≥1.5, 1.67, 2.0, 2.5, etc. for 5-year risk and ≥12, 13, 14, etc. for lifetime risk). The numbers of cases and controls were determined at these risk thresholds, and ORs were calculated at each risk increment to provide a measure of the strength of association of an elevated risk score with breast cancer [54]. The OR was calculated individually for both the PFRM and the BCRAT and plotted. Additionally, positive likelihood ratios (PLRs) were calculated as the proportion of cases at a given risk threshold divided by the proportion of disease-free controls at that threshold [55]. The fold improvement for the PFRM compared to the BCRAT was calculated by dividing their ORs and PLRs, and statistical significance was assessed using the χ^2-test. Area under the receiver operator curve (AUC) analyses were performed as previously described [56]. Model calibration was examined as previously described [15]. The average estimated risk at 5-year intervals was determined for controls ≥35 years in the model building (3176 women) and validation (761 women) sets for the age intervals <45, 46–55, 56–65, and >65 years. The calculated average risks were compared to previously published average risks determined for the same age intervals for 1744 Caucasian control women for the BCRAT (Gail model 2) and a new version of the BCRAT that included breast density and weight [15].

3. Results

3.1. Description of the optimized PFRM

Multivariate logistic model building based on the population of 5022 Caucasian women (1671 breast cancer cases and 3351 cancer-free controls) resulted in an optimized PFRM containing 22 SNPs (from 19 distinct genes), age at first live birth, number of first degree relatives with breast cancer, and previous biopsy number/outcome. Table 1 shows a detailed description of the genes, the SNPs and their major functional pathway. The majority of the SNPs are in genes involved in steroid hormone metabolism or DNA repair. Fig. 1 illustrates how the multiple gene polymorphisms and clinical risk factors are utilized in the PFRM. The ellipses are used to illustrate the role of the terms in the model with terms utilized individually in the left ellipse and those interacting with age in the right ellipse. The SNP terms in the overlapping region of the two ellipses are weighted both individually (Panel A, for all women ages 30–69) as well as by interaction with a specific age stratum (30–49 years) in the absence (Panel B) or presence (Panel C) of having an affected first degree family relative.

3.2. PFRM performance in the model building and validation study populations

PFRM performance was examined in comparison to the BCRAT which, as discussed above, is currently the primary risk evaluation tool applicable to most women. Fig. 2 shows graphs of the ORs determined at increasing risk threshold levels for either 5-year or lifetime absolute risk outputs from the PFRM (solid line), compared to BCRAT (dashed line) and to the random assignment of cases and controls (solid line). Upper panels show the results for the model building sample set while the lower panels show the results for the first independent validation set. These graphs clearly illustrate the improved performance of the PFRM compared to the BCRAT over a broad range of clinically relevant risk ranges. For the PFRM, the OR increases as the risk level outputs by the model increase. In contrast, for BCRAT little or no increase in the OR is found as the risk level outputs by the model increase. The 5-year and lifetime performance characteristics for the PFRM compared to BCRAT are similar for both the model building and validation sample sets. Similar relationships between OR and absolute risk outputs were also observed when considering either 10-year or 15-year time frames (data not shown). These PFRM performance characteristics are what would be expected of an improved prediction model, while the nearly horizontal lines for the BCRAT lie just above what would be observed for a model that randomly assigns risk (i.e. as many cases as controls have elevated risk).

This improvement in risk classification for the PFRM compared to the BCRAT can also be expressed in terms of fold improvement by calculating the ratio of the PFRM-OR/BCRAT-OR. Table 2 shows these data for selected risk thresholds for 5-year and lifetime risks in the model building sample set. For a 5-year risk threshold of ≥1.67% (about the minimum risk for a normal 65 year old Caucasian woman), the PFRM exhibits a 1.3-fold improvement that is statistically significant ($p = 0.002$). Fold-improvements in 5-year risk estimates for the PFRM over the BCRAT of 1.4, 1.9 and 2.7 are seen at risk thresholds of ≥2.0, 3.0, and 4.0%, respectively, that are all statistically significant ($p = 0.002$, $p < 0.0001$, and $p < 0.0001$). Likewise, at least comparable fold improvements are seen in PFRM over BCRAT lifetime risk estimates between 2.6 and 2.8 at risk threshold of ≥20 and 25%, respectively, and these are also highly significant ($p < 0.0001$). The validation set is a smaller sample and as might be expected there is some variability in the OR values compared to the larger model building set (Fig. 2, lower vs. upper panels). Here for the 5-year risks, we observe increased fold improvements of 1.2, 1.2, 1.2, 1.6 and 2.1 at risk thresholds of ≥2.0, 2.5, 3.0 and 3.5%, respectively; these fold improvements trend toward statistical significance at 3.0% risk ($p = 0.07$) and reach significance at the 3.5% risk threshold ($p = 0.037$). In this validation set and at the ≥20% threshold for lifetime risk, a 2.0-fold improvement that is statistically significant ($p = 0.036$) is observed. Thus, despite the much smaller size of the validation sample set, results similar to those from the model building set support the overall conclusion that 5-year and lifetime risk estimations for the PFRM are significantly improved over the BCRAT.

The application of receiver operating characteristic (ROC) curve areas (AUC) to probability-based risk models such as the PFRM and BCRAT has been controversial but persists as a method of evaluation [57]. In our analyses we have found that the new PFRM consistently exhibits improved AUCs compared to the BCRAT. Because of the age stratification used in model building, we performed AUC analysis of 5-year risk outputs for women that fall into 5-year age intervals. The largest improvement in the PFRM compared to BCRAT was in women of 35–39 years of age where the AUC of the PFRM was 0.69 compared to 0.58 for the BCRAT, a difference of 0.11 with a standard error of 0.02, a statistically significant increase (two-sided $p = 0.0002$). Improvements in AUC ranging from 0.06 to 0.14 ($p = 0.02$ to 0.0003) were observed for all but one 5-year age group (age 50–54), where the difference of 0.05 did not reach statistical significance. The average improvement in AUCs for the PFRM compared to BCRAT is similar in the age groups 35–49 (0.08) and 50–69 (0.09).

3.3. PFRM calibration in the model building and validation study populations

The improved performance of the PFRM is the result of the optimized model generating risk estimates that differ considerably from those of the BCRAT. Thus, it is important to show that with this change in risk estimates the improved model remains well calibrated. Table 3 shows a comparison of the average 5-year risks (<45, 46–55, 56–65, >65 years) calculated for our model building and validation sample sets. Comparison of these values obtained for average 5-year risks in our sample sets to published results obtained for the BCRAT (Gail model 2) and the BCRAT-D/W (Gail model 2 with breast density and weight) in their control sample set for different age intervals is shown. These data show that the optimized PFRM is well calibrated, as it produces average risk estimates very close to the very well calibrated BCRAT [8,12], as well as to the more recently examined BCRAT-D/W [15].

3.4. PFRM performance in the high incidence Marin study population

Given that the high incidence Marin County case–control study population has been previously described and shown to have no significant case association with either traditional risk factors or increase in OR assessed by BCRAT [35,36], this study population was compared to our model sample set and PFRM risk scores were determined and compared to BCRAT estimates. As shown in Table 4, and consistent with known demographic differences between the Marin County and other urban US-wide SEER populations [35,36], Marin's cases and controls are balanced with regard to the shown risk characteristics but are significantly older and have later ages at first birth than that of the model building cases and controls. In Table 5, calculated PLRs, based on an elevated lifetime risk threshold of ≥12%, are shown for both PFRM and BCRAT models, comparing the Marin and model building sample sets. Consistent with the previously described OR comparisons between PFRM and BCRAT in the model building population (Table 2), PLRs also demonstrated a significant improvement for PFRM over BCRAT in this study population (1.8 fold improvement, $p < 0.0001$). For the Marin population, in which BCRAT again shows no predictive value (PLR = 0.9), the PFRM yields a PLR = 2.2, representing a significant 2.4 fold predictive improvement over BCRAT ($p = 0.036$). Another measure of improvement in risk estimation is the ability to correctly assign case status at this same threshold level of elevated lifetime risk (≥12%). In the model building

Table 1
Polyfactorial risk model — genes, SNPs and function.

Gene	Gene name	SNP ID — rs#	Base change	SNP location	Function
ACACA	Acetyl coenzyme A carboxylase alpha	rs34915260	T → G	Promoter (PIII)	Fatty acid synthesis and BRCA1 interaction
		N/A	T → C	Exon 1	
		rs2252757	T → C	IVS17	
COMT	Catechol-O-methyltransferase	rs4680	A → G	V158M	Steroid hormone metabolism
CYP11B2	Cytochrome P450, subfamily XIB, polypeptide 2	rs1799998	T → C	Promoter, nt-344	Steroid hormone metabolism
CYP19	Cytochrome P450, family 19, subfamily A, polypeptide 1	rs10046	T → C	3′UTR	Steroid hormone metabolism
CYP1A1	Cytochrome P450, subfamily IA, polypeptide 1	rs4646903	T → C	3′UTR	Steroid hormone metabolism
CYP1B1	Cytochrome P450, subfamily IB, polypeptide 1	rs1800440	A → G	N453S	Steroid hormone metabolism
		rs10012	C → G	R48G	
EPHX	Epoxide hydrolase	rs1051740	T → C	Y113H	Xenobiotic metabolism
ERCC5	Excision repair, complementing defective, in Chinese hamster, 5	rs17655	G → C	D1104H	DNA repair
ESR1	Estrogen receptor 1	rs2077647	T → C	S10S	Steroid hormone metabolism
IGF2	Insulin-like growth factor II	rs2000993	G → A	IVS, nt3580	Growth factor/hormone
INS	Insulin	rs3842752	C → T	nt1107	Growth factor/hormone
KLK10	Kallikrein-related peptidase 10	rs3745535	G → T	A50S	Cell cycle
MSH6	MutS, *E. coli* homolog of, 6	rs3136229	G → A	Promoter, nt-447	DNA repair
RAD51L3	RAD51, *S. cerevisiae*, homolog of, D	rs4796033	G → A	R165Q	DNA repair
SOD2	Superoxide dismutase 2	rs1799725	T → C	V16A	Free radical scavenger
TNFSF6	Tumor necrosis factor ligand superfamily, member 6	rs763110	C → T	nt-844	Apoptosis
VDR	Vitamin D receptor	rs7975232	T → G	Intron 8	Hormone receptor
XPC	Xeroderma pigmentosum, complementation group C	rs2228000	C → T	A499V	DNA repair
XRCC2	X-ray repair, complementing defective, in Chinese hamster, 2	rs3218536	G → A	R188H	DNA repair

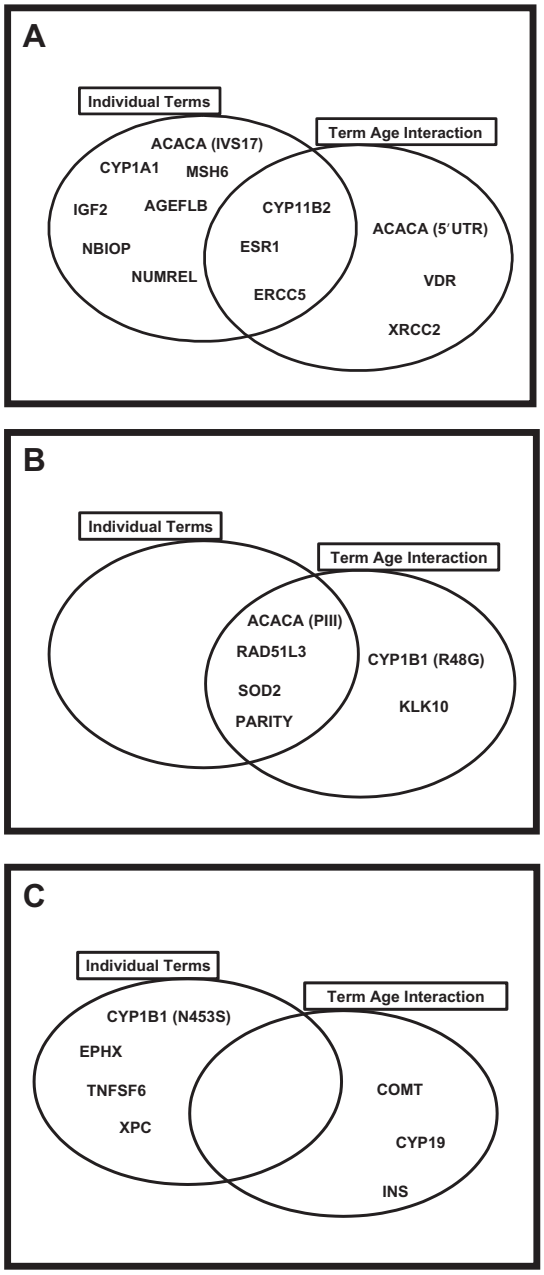

Fig. 1. Illustration of the polyfactorial risk model. In each panel, the left ellipse shows the individual terms in the model and the right ellipse shows the terms interacting with age. The overlapping region in the middle shows terms included both individually and interacting with age. Panel A is for all ages, Panel B is for ages 30–49 without a first degree relative and Panel C is 30–49 with a first degree relative.

sample set, PFRM showed a 27% improvement in identifying case status over BCRAT; in the Marin study sample set, PFRM showed a 51% improvement in identifying case status over BCRAT, again pointing to the improved predictive power of PFRM in populations with increased breast cancer risk.

3.5. A *CYP11B2* variant within PFRM and Marin breast cancer risk

Given that the optimized PFRM incorporates 22 functional SNP candidates with potential cancer relevance, one question arising from the Marin study given the improved performance of PFRM over BCRAT is the relative predictive value of those 22 functional SNPs compared to 7 other well characterized but non-functional GWAS SNPs (rs2981582/*FGFR2*, rs3817198/*LSP1*, rs889312/*MAP3K*, rs4415084/*MRPS30*, rs13281615/*POU5F1P1*, rs13387042/*TNP*, rs3803662/*TOX3*), previously associated with breast cancer risk and also genotyped in the same Marin study. For purposes of comparison with PFRM, a composite GWAS-7/BCRAT risk score was determined by multiplying the estimated risk for each genotype in the Marin population using established ORs and determined allele frequencies. The resulting PLR fold improvement for the GWAS-7/BCRAT lifetime risk estimate over the basic BCRAT risk estimate in the Marin study set was found to be 1.4 (1.1, 5.3 CI), indicating that in this high risk breast cancer population the use of 22 functional SNPs in PFRM with its 2.4 fold improvement over BCRAT offers measureable predictive improvement (2.4/1.4 = 1.7 fold) over the inclusion of 7 GWAS SNPs combined with BCRAT. This observation also led to a search among the 22 SNPs in PFRM for specific functional gene variants most enriched in the Marin population having elevated (≥12%) PFRM risk. Despite the fact that increased Marin breast cancer risk has been shown to be due to an excess annual incidence of estrogen receptor (ER)-positive breast cancers [58], the two PFRM SNPs most closely associated with ER or its metabolism (rs2077647/*ESR1*, rs4680/*COMT*) suggested no genotype enrichment linked to elevated PFRM risk in Marin. However, three other PFRM SNPs functionally associated with steroid hormone metabolism showed some enrichment (in order of magnitude): rs1799998/*CYP11B2*, rs7975232/*VDR*, and rs1800440/*CYP1B1*. A pilot study suggested possible risk association with the vitamin D receptor polymorphism (rs7975232/*VDR*) in this same Marin population [59], but a recent meta-analysis has shown no significant risk association with this specific *VDR* SNP in other breast cancer populations [60]. Attention therefore focused on the PFRM SNP with the greatest genotype enrichment in the higher risk Marin population, the aldosterone synthase gene variant *CYP11B2* (rs1799998), known for its functional and epidemiological link to essential hypertension [61,62] and previously associated with breast cancer risk in an age-specific manner [32]. As summarized in Table 6, while the C/C genotype of this *CYP11B2* promoter variant (-344T/C) is found in 20% of the overall Marin study population, it occurs in 48% of those with elevated PFRM risk, representing a significant 2.4 fold frequency enrichment associated with increased breast cancer risk in Marin ($p < 0.0001$).

4. Discussion

Providing personalized risk assessment for women undergoing breast cancer screening is an increasingly important concern among all health care providers and consumers. If women with a higher lifetime risk of developing breast cancer can be accurately identified relatively early in life, they could be offered more intensive surveillance (e.g. screening by magnetic resonance imaging) as well as prevention measures, both in a more clinically and cost effective manner. As one of the stronger known clinical risk factors, first degree family history of breast cancer is identified in only ~10% of all breast cancer cases, and while genetic risk assessment by screening for *BRCA1* and *BRCA2* mutations can be lifesaving for women with the greatest lifelong susceptibility to developing familial breast and ovarian cancers, the frequency of these highly penetrant germline mutations in the general population is only 0.2% [58,59]. In contrast, all women carry common genetic polymorphisms (SNPs) with variably low penetrant risk connections to sporadic breast cancer development. The BCRAT, which employs only clinical risk factors but has long been used to identify populations that can benefit from prevention protocols [7,8], has also been criticized for lacking the discriminatory accuracy needed for more individualized risk assessment and counseling [12,13]. Unfortunately, efforts to improve the predictive value of BCRAT by incorporating some of the strongest risk-associated GWAS SNPs (e.g. GWAS-7/BCRAT) have achieved only limited success [27–31].

In the present study, we describe the building, performance and calibration of an optimized PFRM that incorporates the same five clinical

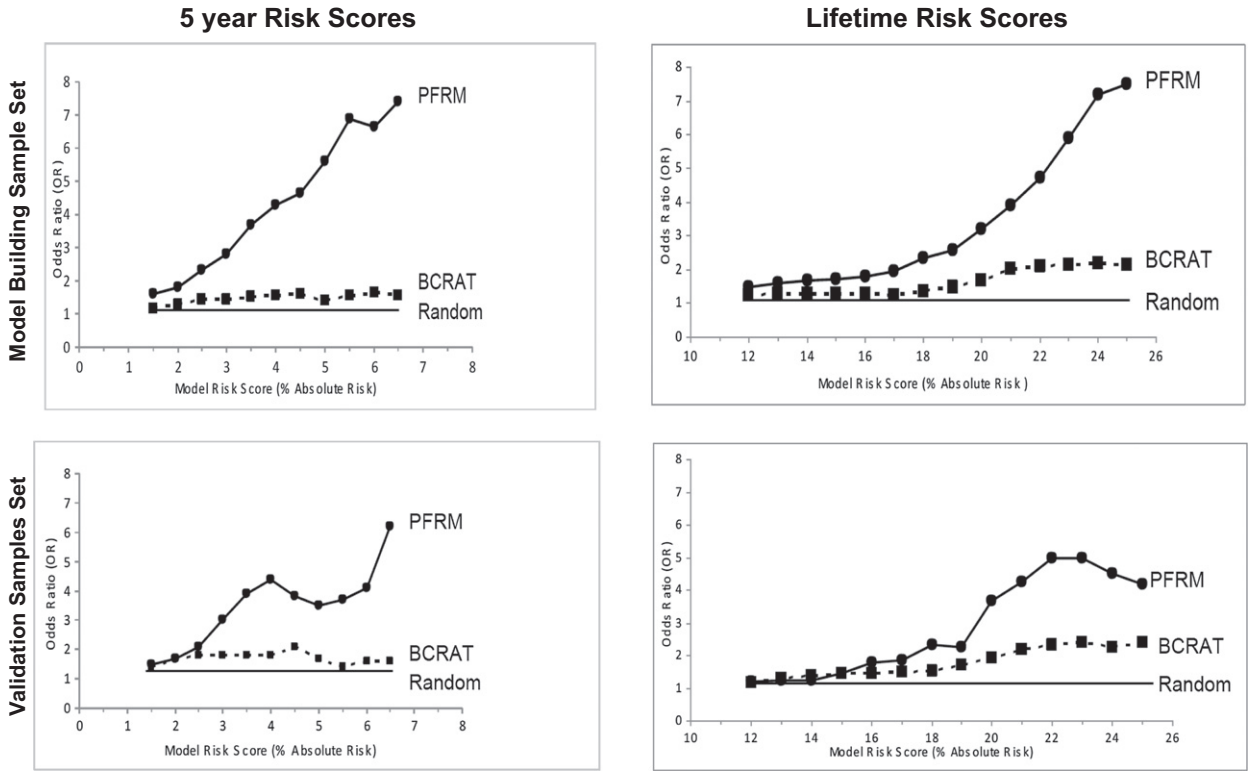

Fig. 2. Odds ratio of breast cancer at increasing model 5-year and lifetime risk scores. Odd ratios (ORs) were calculated at increasing model absolute risk outputs for both the PFRM and BCRAT. The relationships between the OR (y-axis) and risk score (x-axis) are shown for both model building and validation sample sets, and for both 5-year and lifetime risk scores, as indicated. The PFRM is represented by the solid line with circles and the BCRAT is represented by the dotted lines with squares. The solid line at OR = 1 illustrates the line that would be obtained for a model with random assignment of risk scores. The plots are initiated at the mean control population risk obtained from the PFRM.

risk factors used in the BCRAT along with 22 SNPs (in 19 genes) of known functional relevance. This optimized PFRM consistently outperformed the BCRAT in all case–control study sets evaluated, including the model building and two other geographically independent validation populations, totaling 6553 Caucasian women (2236 cases, 4317 controls). Across all three populations, using either ORs or PLRs to estimate risk prediction performance over either 5-year or lifetime risk intervals, the PFRM proved to have at least two fold better predictive performance than BCRAT. At the high lifetime risk threshold of ≥20%, fold improvements in OR for PFRM over BCRAT were 2.6 ($p < 0.0001$) and 2.0 ($p = 0.036$) for the model building and validation sets, respectively, and at the lower lifetime risk threshold of ≥12% (1.5× the SEER average risk over ages 30–69), fold improvements in PLR for PFRM over BCRAT were 1.8 ($p < 0.0001$) and 2.4 ($p = 0.036$) for the model building and Marin populations, respectively. For the high incidence Marin population in particular, at the 12% risk threshold the BCRAT showed virtually no predictive ability (0.9 PLR), unlike the PFRM (2.2 PLR) which proved nearly 2-fold better at assigning case status among the Marin case–control study samples. Although the standard BCRAT, as employed in all these population set comparisons with PFRM, does not incorporate SNPs into its risk estimation, in two preliminary studies we have reported that the PFRM significantly outperforms risk scores obtained using either a core group of 7 GWAS SNPs alone or in combination with BCRAT [63,64]; those latter results obtained using the Marin study population are now presented here.

The full panel of 117 functional SNPs originally included in the PFRM model building was selected based on their biological and physiological plausibility in contributing to the development of breast and related cancers. Based on the importance of estrogens in influencing breast cancer risk and development, perhaps it is not surprising that the majority of the 22 SNPs ending up in the optimized model were genes related to steroid hormone metabolism and DNA repair, with 7 of the SNPs occurring in genes involved in steroid hormone synthesis, signaling or metabolism and 5 within genes related to DNA repair. Curiously, in the Marin population at highest risk for developing ER-positive breast cancer [58], neither *ESR1* (encoding ERα) nor *COMT* (encoding the catechol-O-

Table 2
Fold improvement in ORs at selected risk thresholds.

Risk threshold	OR (95% CI)[a]		Fold improvement (95% CI)	p-Value
	PFRM	BCRAT		
5-Year (%)				
1.67	1.6 (1.4, 1.9)	1.2 (1.1, 1.3)	1.3 (1.1, 1.4)	0.002
≥2.0	1.8 (1.6, 2.1)	1.3 (1.1, 1.5)	1.4 (1.1, 1.6)	0.002
≥3.0	2.8 (2.2, 3.5)	1.4 (1.2, 1.7)	1.9 (1.4, 2.5)	<0.0001
≥4.0	4.3 (3.0, 6.0)	1.5 (1.2, 2.0)	2.7 (1.8, 4.1)	<0.0001
Lifetime (%)				
≥20	4.3 (3.0, 6.0)	1.7 (1.4, 2.0)	2.6 (1.8, 3.8)	<0.0001
≥25	5.6 (3.4, 9.1)	2.0 (1.7, 2.6)	2.8 (1.7, 4.7)	<0.0001

[a] OR = odds ratio and CI = confidence interval.

Table 3
Polyfactorial risk model calibration.

		Average predicted 5-year risk intervals (%)			
		Age ≤ 45	Age 46–55	Age 56–65	Age > 65
Model building	PFRM	0.7	1.4	2.0	2.4
	BCRAT	0.8	1.7	2.4	2.7
Validation	PFRM	0.7	1.4	2.1	2.3
	BCRAT	0.8	1.5	2.4	2.4
Published BCRAT [15]	BCRAT	0.7	1.5	2.1	1.9
	BCRAT-D/W	0.8	1.5	2.2	1.9

Table 4
Comparison of the Marin and model building case–control populations.

	Number of first degree relatives	Marin County		Model building	
		Cases	Controls	Cases	Controls
Mean age[a] (SD)		53.6 (8.17)	54.2 (7.88)	49.7 (9.2)	49.7 (9.3)
Mean age at menarche (SD)		12.7 (1.4)	12.5 (1.4)	12.3 (2.5)	12.5 (2.3)
Mean age at first live birth (SD)		27.5 (5.8)	27.9 (6.1)	24.2 (5.1)	24.0 (5.2)
First degree	0	138 (82)	136 (77)	1308 (78)	2531 (75)
Relatives	1	29 (17)	37 (21)	332 (20)	733 (22)
N (%)	≥2	2 (1)	4 (2)	31 (2)	87 (3)

[a] Age at diagnosis cases and age enrollment controls; SD = standard deviation.

methyltransferase that metabolizes estrogen) demonstrated any significant genotype frequency enrichment associated with increased PFRM risk, although it is important to note that all of the weakly penetrant SNPs in the PFRM contributed in assigning individual risk within the Marin study set that proved collectively more predictive than either the BCRAT or clinical risk factors alone. To be singled out, however, for its contribution to increased PFRM performance in the Marin population is one of the steroid hormone metabolizing SNPs, a promoter variant (-344T/C) of the cytochrome P450 aldosterone synthase gene *CYP11B2*, whose C/C genotype was previously associated with increased breast cancer risk among older age females (55–69 years) [32], but whose same inherited genotype has gained considerable attention among vascular researchers for its strong meta-analysis association with reduced risk of developing essential hypertension in Caucasians [61,62]. Significantly, the *CYP11B2* C/C genotype observed in 20% of the overall Marin study population (consistent with its occurrence in other Caucasian populations) was found in 48% of those with elevated PFRM risk in Marin ($p < 0.0001$).

What connection might a functional gene variant associated with mechanistic regulation of blood pressure via aldosterone synthesis also have in determining breast cancer risk? As it turns out, a very recent and independent study of breast cancer risk in Marin County, conducted by the Marin Women's Study, has uncovered a novel association between the protective effect of pregnancy-induced hypertension (PIH) and later life breast density and breast cancer risk, modified strongly by two specific SNPs (rs3025039/*VEGF*, rs2016347/*IGFR1*) [65]. While neither PIH nor those two *VEGF* and *IGFR1* SNPs were assessed in the Marin case–control study set evaluated here, it is of mechanistic relevance that the same *CYP11B2* promoter SNP linked here to Marin breast cancer risk has previously been implicated (along with VEGF and inappropriate aldosterone production) in the genetic and pathophysiologic basis for the hypertensive pregnancy disorder known as preeclampsia [66–68]. Based on our PFRM findings in the Marin population, future studies should now explore the age- and ethnicity-dependent mechanism potentially connecting inherited aldosterone-altering *CYP11B2* variants with premenopausal predisposition to PIH or preeclampsia, and postmenopausal breast cancer risk.

A major strength of the current study is the presentation of a new PFRM with significantly improved model performance over BCRAT, as validated in two independent study populations while maintaining model calibration. Another strength is inclusion of the Marin County case–control validation study set, with its age and ethnicity matching drawn from a known high incidence breast cancer population. However, a significant limitation of this latter validation study is that recruitment was retrospective and it represents a relatively small study sample. Therefore, observations drawn from this validation study, including the finding of higher breast cancer risk linked to a *CYP11B2* aldosterone synthase SNP, should be considered hypothesis generating and must be tested in another independent, larger, age- and ethnicity-matched population study. Future evaluation of the PFRM in a large, longitudinally followed study cohort is also needed to further validate its general applicability, confirm its calibration, and prospectively test its individualized prediction estimates. Further validation of its improved level of individualized risk prediction could immediately impact clinical practice by more accurately identifying women most likely to benefit from more frequent clinical surveillance, more sensitive screening methods, and/or more aggressive prevention intervention.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbacli.2014.11.001.

Table 5
Fold improvement in PLRs at elevated risk threshold.[a]

Sample set	PLR (95% CI)		Fold improvement (95% CI)	p-Value
	PFRM	BCRAT		
Marin County	2.2 (1.1, 4.3)	0.9 (0.6, 1.3)	2.4 (1.1, 5.65)	0.036
Model building	2.1 (1.8, 2.5)	1.2 (1.1, 1.3)	1.8 (1.4, 2.2)	<0.0001

[a] ≥12%, PLR = positive likelihood ratio, and CI = confidence interval.

Table 6
CYP11B2 genotypes and frequencies in Marin case–control population.

CYP11B2 rs1799998	C/C	C/T	T/T	n
	n (%)	n (%)	n (%)	
PFRM ≥ 12%[a]	16 (48)*	14 (42)	3 (9)	33
Overall study	68 (20)**	167 (49)	103 (30)	338
Cases	28 (17)	89 (54)	47 (29)	164
Controls	40 (23)	78 (45)	56 (32)	174
Elevated risk/all	48%/ 20% = 2.4***	42%/ 49% = 0.86	9%/ 30% = 0.30	

[a] 1.5× the SEER average risk for age range 30–69.
* Mean age (SD) = 56.6 (7.4), 31% ≥ 60.
** Mean age (SD) = 53.9 (8.3), 29% ≥ 60.
*** With a one sample t-test of proportions, $p < 0.0001$, CI = 0.34–0.69.

Acknowledgements

This study was supported by funding from the US Army Breast Cancer Research Program (DAMD17-01-1-0358), Oklahoma Center for the Advancement of Science and Technology—Applied Research Program (AR08.1-030, AR05.1025, AR02.2032, AR01.1-050 and AR99.2-007), American Cancer Society (RPG-97-167-01-MGO), Presbyterian Health Foundation, Oklahoma Life Sciences Fund, Swisher Family Trust to InterGenetics and National Institutes of Health, National Center for Research Resources, General Clinical Research Center (M01RR-1447) to the University of Oklahoma Health Sciences Center, and California Breast Cancer Research Program Community Collaboration Research grants 4AB-14801 and 5BB-1201. We thank Laura Blaylock, Jean Kay, and Cristine Morris of InterGenetics for their technical assistance. We also thank the many clinicians and their patients who voluntarily participated in the studies described herein, including those women who first organized and participated in the Marin community-based Adolescent Risk Factors Study.

References

[1] C. DeSantis, J. Ma, L. Bryan, A. Jemal, Breast cancer statistics, 2013, CA Cancer J. Clin. 64 (2014) 52–62.
[2] D. Saslow, C. Boetes, W. Burke, et al., American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography, CA Cancer J. Clin. 57 (2007) 75–89.
[3] S.M. Domchek, A. Eisen, K. Calzone, et al., Application of breast cancer risk prediction models in clinical practice, J. Clin. Oncol. 21 (2003) 593–601.
[4] A.N. Freedman, D. Seminara, M.H. Gail, et al., Cancer risk prediction models: a workshop on development, evaluation, and application, J. Natl. Cancer Inst. 97 (2005) 715–723.
[5] M.C. Mahoney, T. Bevers, E. Linos, W.C. Willett, Opportunities and strategies for breast cancer prevention through risk reduction, CA Cancer J. Clin. 58 (2008) 347–371.
[6] K. Visvanathan, R.T. Chlebowski, P. Hurley, et al., American Society of Clinical Oncology clinical practice guideline update on the use of pharmacologic interventions including tamoxifen, raloxifene, and aromatase inhibition for breast cancer risk reduction, J. Clin. Oncol. 27 (2009) 3235–3258.
[7] M.H. Gail, L.A. Brinton, D.P. Byar, et al., Projecting individualized probabilities of developing breast cancer for white females who are being examined annually, J. Natl. Cancer Inst. 81 (1989) 1879–1886.
[8] J.P. Costantino, M.H. Gail, D. Pee, et al., Validation studies for models projecting the risk of invasive and total breast cancer incidence, J. Natl. Cancer Inst. 91 (1999) 1541–1548.
[9] M.L. Bondy, E.D. Lustbader, S. Halabi, et al., Validation of a breast cancer risk assessment model in women with a positive family history, J. Natl. Cancer Inst. 86 (1994) 620–625.
[10] D. Spielgelman, G.A. Colditz, D. Hunter, E. Hertzmark, Validation of the Gail et al. model for predicting individual breast cancer risk, J. Natl. Cancer Inst. 86 (1994) 600–607.
[11] B. Fisher, J.P. Costantino, D.L. Wickerham, et al., Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant and Bowel Project P-1 Study, J. Natl. Cancer Inst. 90 (1998) 1371–1388.
[12] B. Rockhill, D. Spiegelman, C. Byrne, D.J. Hunter, G.A. Colditz, Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention, J. Natl. Cancer Inst. 93 (2001) 358–366.
[13] M.L. Bondy, L.A. Newman, Assessing breast cancer risk: evolution of the Gail model, J. Natl. Cancer Inst. 98 (2006) 1172–1173.
[14] C.J. Fabian, B.F. Kimler, C.M. Zalles, et al., Short-term breast cancer prediction by random periareolar fine-needle aspiration cytology and the Gail risk model, J. Natl. Cancer Inst. 92 (2000) 1217–1227.
[15] J. Chen, D. Pee, R. Ayyagari, et al., Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density, J. Natl. Cancer Inst. 98 (2006) 1215–1226.
[16] A.M. Dunning, C.S. Healey, P.D. Pharoah, et al., A systematic review of genetic polymorphisms and breast cancer risk, Cancer Epidemiol. Biomark. Prev. 8 (1999) 843–854.
[17] M.M. de Jong, I.M. Nolte, G.J. te Meerman, et al., Genes other than *BRCA1* and *BRCA2* involved in breast cancer susceptibility, J. Med. Genet. 39 (2002) 225–242.
[18] D.F. Easton, K.A. Pooley, A.M. Dunning, et al., Genome-wide association study identifies novel breast cancer susceptibility loci, Nature 447 (2007) 1087–1093.
[19] D.J. Hunter, P. Kraft, K.B. Jacobs, et al., A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer, Nat. Genet. 39 (2007) 870–874.
[20] S.N. Stacey, A. Manolescu, P. Sulem, et al., Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer, Nat. Genet. 39 (2007) 865–869.
[21] S.N. Stacey, A. Manolescu, P. Sulem, et al., Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer, Nat. Genet. 40 (2008) 703–706.
[22] S. Ahmed, G. Thomas, M. Ghoussaini, et al., Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2, Nat. Genet. 41 (2009) 585–590.
[23] P.D. Pharoah, A. Antoniou, M. Bobrow, R.L. Zimmern, D.F. Easton, B.A.J. Ponder, Polygenic susceptibility to breast cancer and implications for prevention, Nat. Genet. 31 (2002) 33–36.
[24] A.C. Antoniou, P.D. Pharoah, G. McMullan, et al., A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes, Br. J. Cancer 86 (2002) 76–83.
[25] C. Szpirer, J. Szpirer, Mammary cancer susceptibility: human genes and rodent models, Mamm. Genome 18 (2007) 817–831.
[26] C.E. Aston, D.A. Ralph, D.P. Lalo, et al., Oligogenic combinations associated with breast cancer risk in women under 53 years of age, Hum. Genet. 116 (2005) 208–221.
[27] M.H. Gail, Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk, J. Natl. Cancer Inst. 100 (2008) 1037–1041.
[28] M.H. Gail, Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model, J. Natl. Cancer Inst. 101 (2009) 959–963.
[29] P.D.P. Pharoah, A.C. Antoniou, D.F. Easton, B.A.J. Ponder, Polygenes, risk prediction, and targeted prevention of breast cancer, N. Engl. J. Med. 358 (2008) 2796–2803.
[30] S. Wacholder, P. Hartge, R. Prentice, et al., Performance of common genetic variants in breast-cancer risk models, N. Engl. J. Med. 362 (2010) 986–993.
[31] M.E. Mealiffe, R.P. Stokowski, B.K. Rhees, R.L. Prentice, M. Pettinger, D.A. Hinds, Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information, J. Natl. Cancer Inst. 102 (2010) 1618–1627.
[32] D.A. Ralph, L.P. Zhao, C.E. Aston, et al., Age-specific association of steroid hormone pathway gene polymorphisms with breast cancer risk, Cancer 109 (2007) 1940–1948.
[33] S.E. Hankinson, G.A. Colditz, W.C. Willett, Towards an integrated model for breast cancer etiology: the lifelong interplay of genes, lifestyle, and hormones, Breast Cancer Res. 6 (5) (2004) 213–218.
[34] S. Paik, S. Shak, G. Tang, et al., A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer, N. Engl. J. Med. 351 (27) (2004) 2817–2826.
[35] C.A. Clarke, S.L. Glaser, D.W. West, et al., Breast cancer incidence and mortality trends in an affluent population: Marin County, California, USA, 1990–1999, Breast Cancer Res. 4 (2002) R13.
[36] M. Wrensch, T. Chew, G. Farren, et al., Risk factors for breast cancer in a population with high incidence rates, Breast Cancer Res. 5 (2003) R88–R102.
[37] D.L. Hartl, A.G. Clark, Principles of Population Genetics, Sunderland, Masschusetts, Sinauer Associates, Inc., 1997
[38] D. Hosmer, S. Lemeshow, Applied Logic Regression: Second Edition, Wiley, New York, 2002.
[39] A.C. Janssens, Y.S. Aulchenko, S. Elefante, G.J. Borsboom, E.W. Steyerberg, C.M. van Duijn, Predictive testing for complex diseases using multiple genes: fact or fiction? Genet. Med. 8 (2006) 395–400.
[40] Y. Saeys, I. Inza, P. Larrañaga, et al., A review of feature selection techniques in bioinformatics, Bioinformatics 23 (2007) 2507–2517.
[41] B. Halldórsson, V. Bafna, R. Lippert, et al., Optimal haplotype block-free selection of tagging SNPs for genome-wide association studies, Genome Res. 14 (2004) 1633–1640.
[42] N. Long, D. Gianola, G.J.M. Rosa, K.A. Weigel, S. Avendaño, Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers, J. Anim. Breed. Genet. 124 (2007) 377–389.
[43] B. Efron, G. Gong, A leisurely look at the bootstrap, the jackknife, and cross validation, Am. Stat. 37 (1983) 36–48.
[44] P.A. Thompson, P.G. Shields, J.L. Freudenheim, et al., Genetic polymorphisms in catechol-O-methyltransferase, menopausal status, and breast cancer risk, Cancer Res. 58 (1998) 2107–2110.
[45] S. Wedren, T.R. Rudqvist, F. Granath, et al., Catechol-O-methyltransferase gene polymorphism and post-menopausal breast cancer risk, Carcinogenesis 24 (2003) 681–687.
[46] M. Bergman-Jungestrom, M. Gentile, A.C. Lundin, S. Wingren, Association between *CYP17* gene polymorphism and risk of breast cancer in young women, Int. J. Cancer 84 (1999) 350–353.
[47] A.B. Spurdle, J.L. Hopper, G.S. Dite, et al., *CYP17* promoter polymorphism and breast cancer in Australian women under age forty years, J. Natl. Cancer Inst. 92 (2000) 1674–1681.
[48] I. De Vivo, S.E. Hankinson, G.A. Colditz, D.J. Hunter, The progesterone receptor Val660 → Leu polymorphism and breast cancer risk, Breast Cancer Res. 6 (2004) R636–R639.
[49] E.R. Jupe, A.A. Badgett, B.R. Neas, et al., Single nucleotide polymorphism in prohibitin 3′ untranslated region and breast-cancer susceptibility, Lancet 357 (2001) 1588–1589.
[50] S.E. Nelson, M.N. Gould, J.M. Hampton, A. Trentham-Dietz, A case–control study of the HER2 Ile655Val polymorphism in relation to risk of invasive breast cancer, Breast Cancer Res. 7 (2005) R357–R364.
[51] Y. Zhu, H.N. Brown, Y. Zhang, T.R. Holford, T. Zheng, Genotypes and haplotypes of the methyl-CpG-binding domain 2 modify breast cancer risk dependent upon menopausal status, Breast Cancer Res. 7 (2005) R745–R752.
[52] Surveillance, Epidemiology, and End Results (SEER) Program. DevCan database: "SEER 13 Incidence and Mortality, 2000–2002, Follow-back year = 1992, with Kaposi Sarcoma and Mesothelioma". National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch. Released April 2005, based on the November 2004 submission. Underlying mortality data provided by NCHS (www.cdc.gov/nchs). Accessed at www.seer.cancer.gov on April 14, 2011.
[53] J. Benichou, M.H. Gail, Variance calculations and confidence intervals for estimates of the attributable risk based on logistic models, Biometrics 46 (1995) 991–1003.
[54] G. Guyatt, D. Rennie (Eds.), Users' Guide to the Medical Literature: Evidence-based Clinical Practice, American Medical Association Press, Chicago, 2002.
[55] S. McGee, Simplifying likelihood ratios, J. Gen. Intern. Med. 17 (2002) 647–650.
[56] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (1988) 837–845.
[57] M.S. Pepe, H.E. Janes, Gauging the performance of SNPs, biomarkers, and clinical factors for predicting risk of breast cancer, J. Natl. Cancer Inst. 100 (2008) 978–979.
[58] C.C. Benz, C.A. Clarke, D.H. Moore II, Geographic excess of estrogen receptor-positive breast cancer, Cancer Epidemiol. Biomarkers Prev. 12 (2003) 1523–1527.
[59] K.M. Dalessandri, R. Miike, J.K. Wiencke, G. Farren, T.W. Pugh, S. Manjeshwar, D.C. DeFreese, E.R. Jupe, Vitamin D polymorphisms and breast cancer risk in a high-incidence population: a pilot study, J. Am. Coll. Surg. 215 (2012) 652–657.
[60] S. Luo, L. Guo, Y. Li, S. Wang, Vitamin D receptor gene Apal polymorphisms and breast cancer susceptibility: a meta-analysis, Tumour Biol. 35 (2014) 785–790.
[61] S. Sookoian, T.F. Gianotti, C.D. Gonzalez, C.J. Pirola, Association of the C-344T aldosterone synthase gene variant with essential hypertension: a meta-analysis, J. Hypertens. 25 (2007) 5–13.
[62] Z. Hlubocka, M. Jachymova, S. Heller, V. Umnerova, V. Danzig, V. Lanska, K. Horky, A. Linhart, Association of the -344T/C aldosterone synthase gene variant with essential hypertension, Physiol. Res. 58 (2009) 785–792.
[63] E.R. Jupe, T.W. Pugh, N.S. Knowlton, D.C. DeFreese, Breast cancer risk estimation using the OncoVue model compared to combined GWAS single nucleotide polymorphisms, Cancer Res. 69 (2009) 704s [abstract].

[64] K.M. Dalessandri, R. Miike, M.R. Wrensch, et al., Breast cancer risk assessment in the high risk Marin County population using OncoVue compared to SNPs from genome wide association studies, Cancer Res. 69 (2009) 664s [abstract].
[65] L.A. Prebil, R.R. Ereman, M.J. Powell, F. Jamshidian, K. Kerlikowske, J.A. Shepherd, M.S. Hurlbert, C.C. Benz, First pregnancy events and future breast density: modification by age at first pregnancy and specific VEGF and IGF1R gene variants, Cancer Causes Control 25 (2014) 859–868.
[66] G. Escher, M. Cristiano, M. Causevic, M. Baumann, F.J. Frey, D. Surbek, M.G. Mohaupt, High aldosterone-to-renin variants of *CYP11B2* and pregnancy outcome, Nephrol. Dial. Transplant. 24 (2009) 1870–1875.
[67] F.J. Valenzuela, A. Perez-Sepulveda, M.J. Torres, P. Correa, G.M. Repetto, S.E. Illanes, Pathogenesis of preeclampsia: the genetic component, J. Pregnancy 2012 (2012) 632732.
[68] C. Delles, E.M. Freel, Aldosterone, vascular endothelial growth factor, and preeclampsia: a mystery solved? Hypertension 61 (2013) 958–960.