Clinical Validity Assessment of a Breast Cancer Risk Model Combining Genetic and Clinical Information

Matthew E. Mealiffe[1], Renee P. Stokowski[1], Brian K. Rhees[1],

Ross L. Prentice[2], Mary Pettinger[2], David A. Hinds[1,3]*

[1] Perlegen Sciences, Inc.

2021 Stierlin Court

Mountain View, CA 94043

[2] Fred Hutchinson Cancer Research Center

1100 Fairview Avenue N, M3-A410

P.O. Box 19024

Seattle, WA 98109-1024

[3] Current address:

23andMe, Inc.

1390 Shorebird Way

Mountain View, CA 94043

* To whom correspondence should be addressed: dhinds@23andme.com

Tel: 650.938.6300

Fax: 650.938.6305

ABSTRACT

Background: The extent to which common genetic variation can assist in breast cancer (BCa) risk assessment is unclear. We assessed the addition of risk information from a panel of BCa-associated single nucleotide polymorphisms (SNPs) on risk stratification offered by the Gail Model.

Methods: We selected 7 validated SNPs from the literature and genotyped them among white women in a nested case-control study within the Women's Health Initiative Clinical Trial. To model SNP risk, previously published odds ratios were combined multiplicatively. To produce a combined clinical/genetic risk, Gail Model risk estimates were multiplied by combined SNP odds ratios. We assessed classification performance using reclassification tables and receiver operating characteristic (ROC) curves.

Results: The SNP risk score was well calibrated and nearly independent of Gail risk, and the combined predictor was more predictive than either Gail risk or SNP risk alone. In ROC curve analysis, the combined score had an area under the curve (AUC) of 0.594 compared to 0.557 for Gail risk alone. For reclassification with 5-year risk thresholds at 1.5% and 2%, the net reclassification index (NRI) was 0.085 ($Z = 4.3$, $P = 1.0\times10^{-5}$). Focusing on women with Gail 5-year risk of 1.5-2% results in an NRI of 0.195 ($Z = 3.8$, $P = 8.6\times10^{-5}$).

Conclusions: Combining clinical risk factors and validated common genetic risk factors results in improvement in classification of BCa risks in white, postmenopausal women. This may have

implications for informing primary prevention and/or screening strategies. Future research should assess the clinical utility of such strategies.

INTRODUCTION

Breast cancer (BCa) risk has both genetic and environmental influences. Although the discovery of the BRCA1 and BRCA2 genes and a decade of subsequent clinical research has led to substantial positive impact on the health of women affected with the Mendelian cancer predisposition syndromes conferred by mutations in these genes (1-3), the vast majority of BCa genetic risk remains unaccounted for. Building on work suggesting the existence of significant polygenic influences on breast cancer risk (4), recent genome-wide association studies (GWAS) (5-10) and a candidate gene association study (11) have demonstrated that an expanding set of single nucleotide polymorphisms (SNPs) are reproducibly associated with BCa risk in Caucasians and, in some cases, in individuals from other racial/ethnic backgrounds.

As with other complex disorders (12-15), the discovery and validation of these risk SNPs has created an opportunity to explore whether a panel of SNPs can be used to predict disease risk and to assess the clinical relevance of such a panel. In the context of breast cancer, the assessment of disease risk has important clinical implications that can impact decisions about appropriate counseling, screening regimens, and risk reduction strategies (3,16-20). Thus, improvements in risk prediction have the potential to impact clinical care if they are demonstrated to have clinical validity and utility.

For sporadic breast cancer, the Gail Model has been commonly used to produce individual risk estimates. It incorporates individual risk factors including family history (BCa among first-degree relatives), personal reproductive history (age at menarche and at first live birth), and

personal medical history (number of previous breast biopsy examinations and presence of biopsy-confirmed atypical hyperplasia) to estimate personal 5-year and lifetime BCa risk (21). A projected Gail 5-year risk score (e.g. >1.66%) has implications for BCa primary prevention in the context of identifying a group of women who may benefit from risk reduction with a selective estrogen receptor modulator (SERM) (19,20). However, both the discriminatory accuracy of the Gail Model and its calibration in certain populations has been challenged, and uptake of primary prevention strategies amongst physicians caring for women at increased risk for sporadic breast cancer has been modest (22-24).

Two recent papers have evaluated the potential impact of adding genetic information from a panel of 7 breast cancer-associated SNPs to the Gail risk model (25,26). The first analysis predicted, using receiver operating characteristic curves, that area under the curve (AUC) would improve from 0.607 for the Gail model alone to 0.632 with SNP information added to the Gail model (25). In an accompanying editorial, Pepe and Janes suggested that ROC curve analysis might not be particularly useful or relevant in this case and that the use of a reclassification table-based approach could allow the assessment of the fraction of individuals whose risk levels can be meaningfully reclassified (27). The second study showed that real gains—albeit modest—could be realized in reclassification of risk, under the assumption that the combined model was well calibrated (26).

Although several papers have began to set expectations for potential clinical gains from adding a multi-SNP panel to Gail Model risk assessment, they have all been theoretical in nature

(25, 26, 28). Here we show, in a nested case-control study from the Women's Health Initiative (WHI) Clinical Trial (29), that genetic information from a panel of SNPs can be combined with clinical information (i.e., Gail risk) to modestly improve BCa risk estimates in a clinically valid manner in postmenopausal, white women.

METHODS

Description of Case and Control Ascertainment and Sample Handling

We identified all invasive BCa cases that developed between randomization and the originally planned end of the intervention phase of the WHI Clinical Trial (29) that had adequate DNA available for genotyping. We selected one control without a cancer diagnosis for each case, matching on baseline age, self-reported ethnicity, clinical trial participation, years since randomization, and hysterectomy status. Written informed consent was obtained from each woman for her WHI participation. Human investigations were performed after approval by the Fred Hutchinson Cancer Research Center Institutional Review Board. Research subjects had an opportunity to opt in or out of any collaborations involving commercial entities. We restricted our analyses to the subset of these individuals that had consented for commercial use. Of the individuals in the trial, 84 percent provided such consent. The interventions used in the WHI clinical trial are independent of baseline genetic and clinical risk factors by study design, so analyses presented here were not stratified by trial intervention. Due to availability of accurate SNP odds ratios and also Gail model validity, we focused our analyses in this work on non-

Hispanic white women in this nested case-control study, representing 87% of the matched cases and controls. Clinical characteristics of these subjects are summarized in Table 1.

Gail Model and SNP Selection

We used the Gail Model to estimate 5-year risk of BCa based on age, ethnicity, age at menarche, age at first live birth, number of first degree relatives with BCa, and number of previous breast biopsies. We did not have information on biopsy histopathology (i.e., whether atypical hyperplasia was present), so this was coded as "unknown." We scored the subjects using a re-implementation of the current Breast Cancer Risk Assessment Tool (BCRAT) risk calculator from source code downloaded from the National Cancer Institute website (07-Aug-2008).

We selected SNPs to include in the risk classifier that had discovery *P* values $< 5\times10^{-7}$ for SNPs with demonstrated association in genome-wide association studies, to account for multiple hypothesis testing and replication in an independent population. Seven SNPs were found to meet these criteria at the time that the study was initiated (5-8). The 7 SNPs selected from peer-reviewed GWAS results have also been reported to be associated with breast cancer with high statistical significance across multiple large sample sets (Table 2).

The Composite SNP Risk Score

SNPs were genotyped via a custom chip and/or on the Sequenom platform (Supplemental Methods). To model SNP risk using the 7 selected breast cancer-associated SNPs, we used the

estimates of per-allele odds ratios previously reported, as shown in Table 2. We used a multiplicative model for odds ratio across SNPs, where risk values for each SNP were scaled to have a population average of 1 based on the expected frequencies of the three possible diploid genotypes. Missing genotypes were also assigned a relative risk of 1. None of the model parameters were estimated using data derived from the WHI cohort.

Statistical Analysis

We used logistic regression to estimate odds ratios associated with log-transformed Gail model score and the composite SNP score, separately and combined. The intercept term in the logistic regression was unrestricted, and allows the analysis to adapt to case-control sampling. The Hosmer-Lemeshow test (30) is typically used to assess model calibration in cohort data. To generate expected numbers of cases and controls, we used logistic regression to refit the intercept for each risk model, holding the coefficient of the log transformed risk score fixed at 1. This rescales the risk scores to match the actual numbers of cases and controls.

We assessed classification performance using receiver operating characteristic (ROC) curves. We used bootstrap resampling (1000 replicates) to estimate confidence intervals for area under the curve (AUC) as well as differences in AUC. We also evaluated classification accuracy using reclassification tables (31,32) and quantified differences in classification by "net reclassification improvement" (NRI) (33). NRI is the sum of proportions of cases moving to a higher risk category minus cases moving to a lower risk category, and proportions of controls

moving to a lower risk category minus controls moving to a higher risk category (33). We again used bootstrap resampling to evaluate confidence intervals for NRI and to determine empirical *P* values for differences in NRI. We used the same bootstrap samples for each classifier and used paired tests to compare classification performance, to preserve the correlation structure of the classifiers and obtain the most powerful tests.

RESULTS

Characteristics of the Women in the Nested Case-Control Cohort

Table 1 presents the summary characteristics of the case and control groups. As expected, several known BCa clinical risk factors had a differential representation in the two groups: age at menarche (P_{trend} = 0.02), age at birth of first child (P_{trend} = 0.004), first degree relatives with BCa (P_{trend} = 0.0001), and number of previous breast biopsies ($P_{trend} = 1\times10^{-5}$).

Individual SNP Associations with Breast Cancer

Six of the seven SNPs tested were included on the custom SNP array for the parent research project. As detailed elsewhere (Y. Huang, D.G. Ballinger, U. Peters, D.A. Hinds, D.R. Cox, E. Beilharz *et al.*, submitted), all six gave per allele odds ratios in the project dataset that were consistent with those used here from the literature (Table 2). The association was significant, with $P < 0.05$, for five of the six, while the sixth (rs13281615) yielded $P = 0.06$. The seventh SNP, rs13387042 was genotyped specifically for this report. The per allele odds ratio

estimate for this SNP was 1.16, with 95% confidence interval from 1.05 to 1.29 (P = 0.003). Hence, there is agreement between SNP associations observed in this cohort and those based on studies in other populations. We did not detect any significant pairwise interactions among the 7 SNPs (21 distinct tests yielded one test with $P < 0.05$ and none with $P < 0.01$). However, this study was only powered to detect relatively strong effects.

Independence and Calibration of the SNP Risk Score

We separately tested Gail 5-year absolute risk and the 7-SNP odds ratio estimate for association with BCa incidence by logistic regression with log-transformed predictors. Both were strongly associated (Table 3). A two-fold increase in Gail risk yields a less-than-two-fold increase in cancer incidence in this cohort, suggesting that Gail risk is not so well calibrated in this dataset, which is consistent with a previous cohort study report from the WHI observational study (23). The multiplicative model for SNP risk gives relative risk estimates roughly proportional to observed disease rates. Gail risk and SNP risk were weakly but significantly correlated (r = 0.042, P = 0.02). A combined predictor formed by multiplying the Gail absolute risk by the SNP relative risk was more strongly associated with BCa risk than either component alone. Including log-transformed Gail risk and SNP risk in the model as separate terms in the logistic regression further improves the fit ($P = 2.3\times10^{-5}$), by accommodating the difference in calibration of the two terms. However, in a model with these separate terms, an interaction term did not further improve prediction of breast cancer status (P = 0.5).

To visualize the effect of SNP risk for different Gail risk categories in a more clinically intuitive way, we binned the scores into quintiles and evaluated the relationship between SNP risk and BCa odds within each Gail risk stratum (Figure 1). SNP risk is consistently associated with BCa within each stratum. This provides additional evidence that the scores are providing essentially independent information about risk.

To assess calibration of our risk scores, we used the Hosmer-Lemeshow test (30), which compares expected and observed counts of cases and controls in deciles of risk (Supplemental Table 2). As shown in Figure 2, our SNP risk scores appear to be well calibrated ($P = 0.12$). Gail risk, however, was not well calibrated ($P = 6\times10^{-7}$), and this lack of calibration carries over to the combined risk score to an intermediate extent ($P = 0.001$). These results are consistent with our logistic regression results. We did not observe an improvement in calibration in the subset of women with no missing observations for Gail risk factors.

We used linear regression to test whether the log-transformed SNP risk score was predictive of any of the clinical risk factors contributing to the Gail model, while adjusting for case/control status. SNP risk was not significantly correlated with age at menarche ($P = 0.96$) or age at menopause ($P = 0.78$), number of first degree relatives with BCa ($P = 0.20$) or number of previous breast biopsies ($P = 0.41$). The SNP risk score was most associated with age at first birth ($P = 0.10$), and this association appeared to be specifically mediated by rs2981582 in *FGFR2*. This SNP alone gave a stronger signal ($P = 0.008$) and was the only SNP associated with any Gail clinical component at better than nominal $P < 0.05$. Considering only controls, age

at first birth gave $P = 0.03$ for association with SNP risk and $P = 0.002$ for association with rs2981582, and there were no other associations with nominal $P < 0.05$. Given the modest P values, these results will require replication and validation in other datasets.

Classification Performance of the Combined Risk Score

We assessed classification performance using receiver operating characteristic (ROC) curves. We compared classification using 5-year Gail absolute risk, SNP risk, and the combined SNP×Gail risk score (Figure 3). The combination of Gail and SNP risk had an AUC of 0.594 (95% CI: 0.575 – 0.612), compared to 0.557 (95% CI: 0.537 – 0.575) for Gail risk alone and 0.587 (0.567 – 0.607) for SNP risk alone. The difference in AUC for the combined risk score versus Gail alone was statistically significant (95% CI: 0.025 – 0.051, empirical $P < 0.001$).

We also evaluated classification accuracy using reclassification tables (30,31) and quantified differences in classification by "net reclassification improvement" (NRI) (32). While reclassification tables are most clinically intuitive in the context of population-based cohort studies, they still provide useful information here. Specifically, the NRI metric is unaffected by case control sampling. We chose 5-year risk thresholds of 1.5% (for below-average risk) and 2% (for elevated risk) and evaluated reclassification for the combined SNP×Gail score versus Gail risk alone (Table 4). The combined risk score tends to push individual risks towards the tails of the risk distribution: it places 22% fewer cases and 29% fewer controls in the intermediate 1.5-2.0% bin compared to the Gail score alone. The NRI for this table is 0.085 ($Z = 4.3$, $P = 1.0\times10^{-5}$), indicating that these

changes also tend to be in the right direction. Classification improved for 5.6% of cases (P = 4.8×10^{-5}) and 2.9% of controls (P = 0.018).

Reclassification performance is related to the number of clinically meaningful risk categories, because binning of risk scores conceals improvements in risk estimates that do not cross the prespecified thresholds. If risk thresholds are chosen to split women into quintiles of Gail risk (breaks at 1.2%, 1.5%, 1.8%, and 2.4%), then NRI increases to 0.141 (Z = 5.63, P = 9.0×10^{-9}). For deciles of Gail risk, NRI is 0.182 (Z = 6.2, P = 2.1×10^{-10}).

The cost effectiveness of a genetic test can be improved by avoiding testing individuals whose status is unlikely to change as a result of the test. Individuals who are far from the classification cut points are unlikely to be reclassified as a result of the test, and as a result, it is less efficient to test them. This effect should be reflected in NRI, because excluding the tails of the risk distribution should result in reclassification of a higher proportion of tested individuals. We evaluated NRI across a grid of possible lower and upper bounds of Gail risk. Excluding women in the tails of the risk distribution resulted in higher NRI (Supplemental Figure 2). If all women with Gail risk < 1.5% or >2.0% are excluded, then NRI improves to 0.195 (Z = 3.8, P = 8.6×10^{-5}).

Performance in Women with Previous Breast Biopsies

Because women with previous breast biopsies are a group at intermediately elevated risk of BCa in particular need of risk stratification to guide future screening and preventative strategies, we also assessed the impact of the SNP risk score in this subset. Not surprisingly given the loss

of an important risk stratifier, the Gail model had an AUC of only 0.514 (95% CI: 0.471 – 0.561) in this subset. The combined model had an AUC of 0.571 (95% CI: 0.526 – 0.614). We also computed reclassification metrics in the biopsy subset (Supplemental Table 3). In this subset, the NRI is 0.175, which is significant despite the smaller number of events ($Z = 3.9$, $P = 4.9\times10^{-5}$). Here, classification improved for 14.8% of controls ($P = 1.5\times10^{-5}$) but only 2.8% of cases ($P = 0.16$). We used bootstrap resampling to evaluate whether the difference in NRI between the full cohort and the biopsy subset was statistically significant. Based on 1000 bootstrap replicates, a 95% confidence interval for the improvement in NRI in the biopsy subset extended from 0.02 to 0.16, with empirical $P = 0.03$. This increase in NRI is not simply a consequence of conditioning on an important Gail risk factor: in the subset of individuals without a previous breast biopsy, NRI is reduced to 0.065.

DISCUSSION

Recently, several GWAS have demonstrated a number of distinct loci and SNPs that are convincingly associated with risk for sporadic breast cancer. Clinical experience with highly penetrant Mendelian breast cancer risk syndromes and trials of breast cancer risk reduction in individuals at modestly elevated risk, suggest that better risk assessment lays the groundwork for clinical improvements in surveillance and risk reduction strategies (3,17). Thus, there has been substantial interest in assessing the potential impact of combined BCa SNP risk panels on BCa risk assessment (25-28).

Here we have assessed the risk prediction performance of a panel of 7 validated breast cancer risk SNPs in the context of a nested case-control study from the WHI Clinical Trial. Although there was a statistically significant weak correlation between the Gail risk score and SNP risk, the two are essentially independent for practical purposes in risk prediction. We also assessed calibration of SNP and Gail risks in this cohort. Notably, SNP risk score calibration, as measured by the Hosmer-Lemeshow test, was good. Thus, our results support a simple multiplicative model for combining SNP risks.

The calibration and discrimination of the Gail model in the WHI cohort is known to be somewhat worse than has been seen in other large studies (23). This is likely due to a combination of factors, including higher mammography rates, differences in age distributions, and changes in breast biopsy procedures, with more common use of image-guided percutaneous core biopsy procedures which have a lower threshold for use than open biopsy (23). The lack of data on atypical hyperplasia may have also contributed somewhat. Additionally, as the WHI clinical trial tested the impact of hormone replacement therapy (HRT) on breast cancer risk, a higher percentage of women in our case-control cohort may have received HRT than in the studies in which the Gail model was previously validated. Lastly, the age matching in our nested case-control study is likely also contributing significantly. This does not directly impact our assessment of independence and calibration of the SNP risk scores, but may affect the quantitative metrics of improvement in risk assessment in the combined model. Reclassification performance is

sensitive to model calibration as well as discrimination, and will need to be further characterized in population based cohorts.

When we assessed the performance of a combined risk predictor incorporating both the Gail risk and SNP risk, the combined risk predictor performed better in predicting BCa risk than either Gail risk or SNP risk alone. By ROC curve analysis, the AUC for Gail plus SNP risk was 0.594 (95% CI: 0.575 – 0.612) as compared to 0.557 (95% CI: 0.537 – 0.575) for Gail risk alone. Although this statistically significant improvement is modest, in this dataset the Gail model itself had an AUC that was only 5.7% greater than that expected by chance.

Although ROC curves are useful in some contexts, they have been criticized for several reasons: (1) they summarize classification information across the full range of sensitivities and specificities (in most clinical contexts, only a subset of these sensitivities and specificities are relevant); (2) they do not provide information about the actual risks predicted by the model; (3) they do not provide information about the proportion of individuals with particularly high or low risk values; and (4) the area under the ROC curve, which is the probability that a predicted risk for an individual with an event is higher than for an individual without an event, has minimal direct clinical relevance (34). Therefore, we utilized reclassification tables (30,31) to calculate a net reclassification improvement (32) which is a more helpful measure of the potential impact of the combined Gail plus SNP test (27,31-32,34). NRI is a relatively new statistic, but has gained increasing acceptance as an important part of the evaluation of new biomarkers and risk stratifiers (34, 35). This analysis demonstrated a statistically significant improvement in classification (NRI =

0.085; $P = 1.0\times10^{-5}$). Importantly, NRI can be substantially improved by focusing SNP genotyping on those individuals who are predicted to be at intermediate risk by the Gail score, as women at intermediate risk are most likely to be reclassified after the addition of SNP risk. For example, limiting SNP testing to women with Gail 5-year risks between 1.5 and 2.0 percent results in an NRI of 0.195. Taking this information into account, future efforts should rigorously evaluate the clinical utility of targeted strategies to incorporate the combined risk score into the clinical decision-making process in the context of both BCa primary prevention and screening (17-20).

We also evaluated whether it might be possible to obtain better test performance by focusing on a subset of women at particularly high risk—those with previous breast biopsies (36,37). Although this analysis was limited by the lack of pathology results to allow identification of those with more serious histopathologies such as atypical hyperplasia, we pursued similar ROC curve and reclassification table based analyses. Here, classification improved for 14.8% of controls ($P = 4.4\times10^{-5}$) but only 2.8% of cases. Although this suggests that the combined SNP plus Gail test might assist in identifying a subset of women with prior biopsies who might not need as aggressive risk reduction and surveillance efforts as their biopsy history suggests, these results should be interpreted with caution and will require further study in other datasets with available histopathology from the previous biopsies.

This study has several strengths. First, we have taken a rigorous approach to identifying the SNPs to include in the panel, only including those that have been reproducibly associated and for which consistent risk estimates have been reported in the literature. Second, we have

genotyped individuals from a large prospectively recruited cohort with meticulous data collection and rigorous ascertainment of relevant BCa outcomes. Third, we have used literature based genetic risk estimates and have combined them in a straightforward fashion to form risk predictors. Importantly, we did not train our predictors on the WHI data and only used WHI samples to assess their performance

Limitations to this study include the composition of the WHI cohort — which limits inferential scope to white, postmenopausal women — and the clinical characteristics of the women therein. For example, individuals within WHI received HRT at a higher frequency than women currently do at present. Importantly, the age matching design inherent to this study does remove one of the Gail model variables and contributes both to the relatively low AUC seen for the Gail model in our analysis and to its poor calibration. In addition, the absence of pathology records for previous breast biopsies in WHI required us to estimate individual Gail risks by coding atypical hyperplasia status as unknown for women with prior breast biopsies. Although the frequency of atypical hyperplasia-containing biopsies is low enough that this seems unlikely to have affected the analyses of the entire nested case-control study, it is unclear to what extent this may have impacted the analyses focusing on the subset of women with one or more previous biopsies. Finally, case control sampling means we cannot evaluate calibration of absolute event rates; we can effectively only test the slope of the relationship between expected and observed risk, and not the intercept. Therefore, while the WHI cohort is sufficient to support the validity of the SNP and combined risk models in predicting BCa risk, there is a need for further assessment

of the clinical validity of the combined Gail model and SNP panel, especially in population-based cohorts.

The major finding from this study is the demonstration that genetic risk information may be combined multiplicatively with Gail risk scores to improve BCa risk estimation in postmenopausal white women. This finding is based both on the observation that patient BCa risk may be accurately estimated by combining published SNP risk estimates, and also the observation that correlation between SNP risk scores and Gail scores for individuals is weak, allowing patient BCa risk to be more accurately estimated by combining SNP and Gail risks multiplicatively. Thus, the present study supports the claim that the combined risk estimation model approach has clinical validity in the broad sense in postmenopausal, white women.

Our analysis has not addressed clinical utility. The use of improved risk models, such as the one described here, may benefit the public health if shown to have clinical utility when combined with optimal individualized screening and risk reduction strategies. A previous evaluation of utility considered an unselective “all comers” strategy for SNP testing (26). The results of our analysis suggest that, as utility is sensitive to how a test is targeted, it may be wise to focus the application of SNP genotyping for breast cancer risk on women at intermediate risk as measured by the Gail model. Such a strategy clearly boosts reclassification performance in this study. Future research should assess performance in population-based cohorts and ultimately take the next step and address whether reclassification improvement can be translated into improved prevention and/or screening outcomes in the clinic.

REFERENCES

1. Miki Y, Swensen J, Shattuck-Eidens D, *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994;266:66-71.

2. Wooster R, Bignell G, Lancaster J, *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* 1995;378:789-92.

3. Robson M, Offit K. Clinical Practice. Management of an inherited predisposition to breast cancer. *N Engl J Med* 2007;357:154-62.

4. Pharoah PDP, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BAJ. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet* 2002;31:33-6.

5. Easton DF, Pooley KA, Dunning AM, *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;447:1087-93.

6. Hunter DJ, Kraft P, Jacobs KB, *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 2007;39:870-4.

7. Stacey SN, Manolescu A, Sulem P, *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2007;39:865-9.

8. Stacey SN, Manolescu A, Sulem P, *et al.* Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 2008;40:703-6.

9. Ahmed S, Thomas G, Ghoussaini M, *et al.* Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* 2009;41:585-90.

10. Thomas G, Jacobs KB, Kraft P, *et al.* A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* 2009;41:579-84.

11. Cox A, Dunning AM, Garcia-Closas M, *et al.* A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* 2007;39:352-8.

12. Zheng SL, Sun J, Wiklund F, *et al.* Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 2008; 358:910-9.

13. Nam RK, Zhang WW, Trachtenberg J, *et al.* Utility of incorporating genetic variants for the early detection of prostate cancer. *Clin Cancer Res* 2009; 15:1787-93.

14. Weersma RK, Stokkers PC, Cleynen I, *et al.* Confirmation of multiple Crohn's disease susceptibility loci in a large Dutch-Belgian cohort. *Am J Gastroenterol* 2009; 104:630-8.

15. Kathiresan S, Melander O, Anevski D, *et al.* Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 2008; 358:1240-9.

16. Domchek SM, Eisen A, Calzone K, Stopfer J, Blackwood A, Weber BL. Application of breast cancer risk prediction models in clinical practice. *J Clin Oncol* 2003; 21:593-601.

17. Mahoney MC, Bevers T, Linos E, Willett WC. Opportunities and strategies for breast cancer prevention through risk reduction. *CA Cancer J Clin* 2008;58:347-71.

18. Saslow D, Boetes C, Burke W, *et al.* American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007;57:75-89.

19. Visvanathan K, Chlebowski RT, Hurley P, *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* 2009; DOI: 10.1200/JCO.2008.20.5179

20. Kinsinger LA, Harris, R, Lewis C, Woddell M. Chemoprevention of breast cancer: A review of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2002;137(1):59-69.

21. Gail MH, Brinton LA, Byar DP, 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.

22. Elmore JG, Fletcher SW. The risk of cancer risk prediction: "What is my risk of getting breast cancer." *J Natl Cancer Inst* 2006; 98:1673-1675.

23. Chlebowski RT, Anderson GL, Lane DS, *et al*. Predicting risk of breast cancer in postmenopausal women by hormone receptor status. *J Natl Cancer Inst* 2007;99:1695-705.

24. Armstrong K, Quistberg DA, Micco E, *et al.* Prescription of tamoxifen for breast cancer prevention by primary care physicians. *Arch Intern Med* 2006; 166:2260-5.

25. Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. *J Natl Cancer Inst* 2008;100:1037-41.

26. Gail MH. Value of adding single-nucleotide polymorphism genotypes to a breast cancer risk model. *J Natl Cancer Inst* 2009;101:959-63.

27. Pepe MS, Janes HE. Gauging the performance of SNPs, biomarkers, and clinical factors for predicting risk of breast cancer. *J Natl Cancer Inst* 2008;100:978-9.

28. Pharoah PDP, Antoniou AC, Easton DF, Ponder BAJ. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med* 2008;358:2796-803.

29. Prentice RL, Anderson GL. The Women's Health Initiative: lessons learned. *Annu Rev Public Health* 2007;29:131-50.

30. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: Wiley; 1989: Section 5.2.2.

31. Cook NR, Buring JE, Ridker PM. The effect of including C-reactive protein in cardiovascular risk prediction models for women. *Ann Intern Med* 2006;145:21-9.

32. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928-35.

33. Pencina MJ, D'Agostino Sr RB, D'Agostino Jr RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2007;27:157-72.

34. Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. *Ann Intern Med* 2008;149:751-60.

35. McGeechan K, Macaskill P, Irwig L, Liew G, Wong TY. Assessing new biomarkers and predictive models for use in clinical practice: a clinician's guide. *Arch Intern Med* 2008;168: 2304-10.

36. Hartmann LC, Sellers TA, Frost MH, *et al.* Benign breast disease and the risk of breast cancer. *N Engl J Med* 2005; 353:229-37.

37. Ashbeck EL, Rosenberg RD, Stauber PM, Key CR. Benign breast biopsy diagnosis and subsequent risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2007; 16:467-72.

Funding

The work of R. Prentice and M. Pettinger was supported by the National Heart Lung and Blood Institute (NHLBI, contract HHSN268200764314C) and the National Cancer Institute (grant PO1 CA53996) at the National Institutes of Health. The Women's Health Initiative is supported by NHLBI contracts.

Notes

M. Mealiffe, R. Stokowski, B. Rhees, and D. Hinds were employees of Perlegen Sciences during this work, and these analyses were relevant to development of a diagnostic product by Perlegen.

We thank Ellen Beasley, Allison Kurian, Kevin Hughes, Anne-Renee Hartman, and Bryan Walser for helpful comments and discussions. Additionally, we thank the anonymous reviewers, whose comments and suggestions meaningfully improved this manuscript. The authors would like to thank WHI investigators and staff for their dedication and the study participants for making the WHI possible. A listing of WHI investigators can be found at http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf.

Table 1. Clinical characteristics of the study participants.

	Controls		Cases		
	N	%	N	%	P_{trend}
Participants	1636	100.0	1664	100.0	
Estrogen receptor (ER) status					
ER-positive tumor			1218	73.2	
ER-negative tumor			208	12.5	
Unknown/missing			238	14.3	
Age at screening					0.84*
50-59	510	31.2	510	30.6	
60-69	767	46.9	789	47.4	
70-79	359	21.9	365	21.9	
Missing	0	0.0	0	0.0	
Age at menarche					0.02
< 12	349	21.3	379	22.8	
12	407	24.9	441	26.5	
13	481	29.4	497	29.9	
≥ 14	395	24.1	338	20.3	
Missing	4	0.2	9	0.5	
Age at birth of first child					0.004
< 20	214	13.1	182	10.9	
20-24	665	40.6	625	37.6	
25-29	349	21.3	370	22.2	
≥ 30	117	7.2	151	9.1	
No term pregnancy	166	10.1	200	12.0	
Missing	125	7.6	136	8.2	
Age at menopause					0.52
< 45	294	18.0	300	18.0	
45-49	390	23.8	355	21.3	
50-54	596	36.4	646	38.8	
> 54	237	14.5	238	14.3	
Missing	119	7.3	125	7.5	
First degree relatives with breast cancer					0.0001
0	1319	80.6	1253	75.3	
1+	226	13.8	309	18.6	

Missing	91	5.6	102	6.1	
Number of previous breast biopsies					1×10^{-5}
0	1148	70.2	1056	63.5	
1	188	11.5	298	17.9	
2+	83	5.1	102	6.1	
Missing	217	13.3	208	12.5	

* Cases and controls were age-matched.

Table 2. Replicated loci associated with invasive breast cancer.

dbSNP rsID	Gene	Location	Freq*	OR (95% CI)†	Reference
rs2981582	FGFR2	10q	0.38	1.26 (1.23 – 1.30)	Easton et al., 2007
rs3803662	TNRC9	16q	0.25	1.20 (1.16 – 1.24)	Easton et al., 2007
rs889312	MAP3K1	5q	0.28	1.13 (1.10 – 1.16)	Easton et al., 2007
rs13387042	(none)	2q35	0.50	1.20 (1.14 – 1.26)	Stacey et al., 2007
rs13281615	(none)	8q24	0.40	1.08 (1.05 – 1.11)	Easton et al., 2007
rs4415084	FGF10	5p	0.44	1.16 (1.10 – 1.21)	Stacey et al., 2008
rs3817198	LSP1	11p	0.30	1.07 (1.04 – 1.11)	Easton et al., 2007

* Frequency of the high risk allele, in the cited study.

† Odds ratio (and confidence interval) per copy of the high risk allele, in the cited study.

Table 3. Logistic regression tests of association with invasive breast cancer.

Predictor	(95% CI)*	OR per 2× risk (95% CI)†	*P*
log(Gail 5-year risk)	0.46 (0.30 – 0.63)	1.38 (1.23 – 1.54)	1.8×10^{-8}
log(SNP risk)	1.11 (0.86 – 1.36)	2.16 (1.82 – 2.57)	6.4×10^{-19}
log(SNP × Gail risk)	0.65 (0.51 – 0.78)	1.57 (1.42 – 1.72)	3.3×10^{-21}

* Logistic regression coefficient for the risk score.

† Odds ratio corresponding to a 2-fold increase in the risk score, equal to 2 .

Table 4. Reclassification table for SNP×Gail risk versus Gail risk.

	SNP risk× Gail5-year risk			
Gail5-year risk	< 1.5%	1.5% -2.0%	> 2.0%	Total
< 1.5%				
Women	1060	241	64	1365
Events	455	133	41	629
Nonevents	605	108	23	736
Proportion	0.429	0.552	0.641	0.461
1.5% -2.0%				
Women	351	342	263	956
Events	155	172	157	484
Nonevents	196	170	106	472
Proportion	0.442	0.503	0.597	0.506
> 2.0%				
Women	43	129	807	979
Events	19	64	468	551
Nonevents	24	65	339	428
Proportion	0.442	0.496	0.580	0.563
Total				
Women	1454	712	1134	3300
Events	629	369	666	1664
Nonevents	825	343	468	1636
Proportion	0.433	0.518	0.587	0.504

Figure Legends

Figure 1. Relationship between odds of breast cancer in the study data, and single nucleotide polymorphism (SNP) risk quintile, stratified by Gail risk quintile. SNP risk is consistently related to breast cancer odds across Gail risk strata.

Figure 2. Observed versus expected event rates, for deciles of (A) the single nucleotide polymorphism (SNP) risk score, (B) Gail 5-year risk, and (C) the combined risk score. If the risk scores are calibrated, then the points should fall along the dashed line with a slope of 1.

Figure 3. Receiver operating characteristic (ROC) curves for Gail 5-year risk, single nucleotide polymorphism (SNP) risk, and combined risk.

Figure 1.

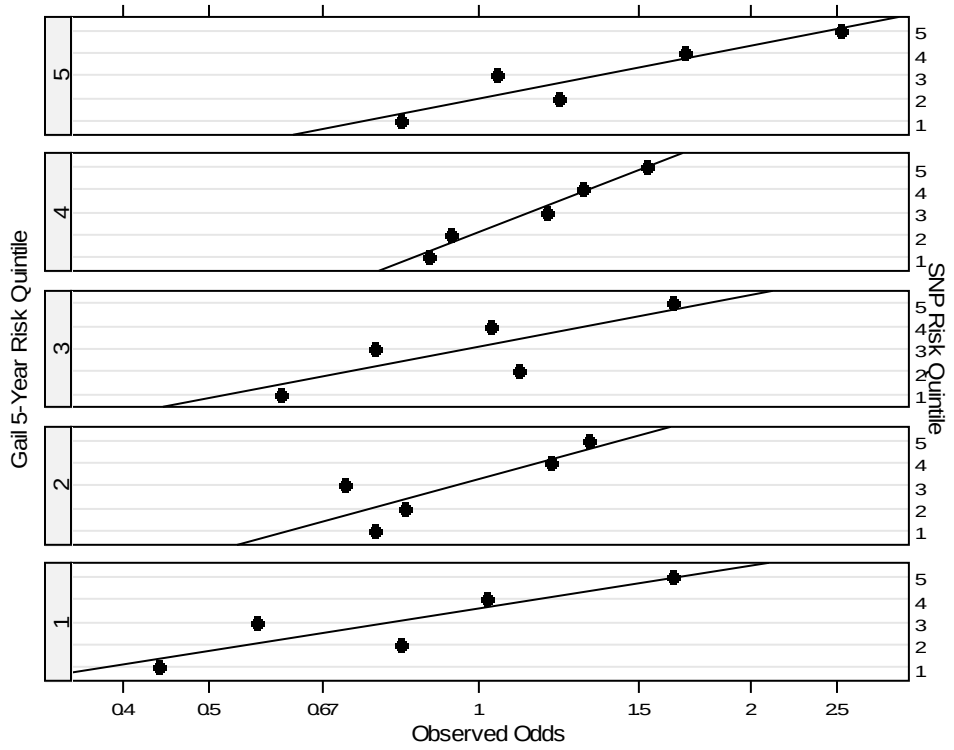


Figure 2.

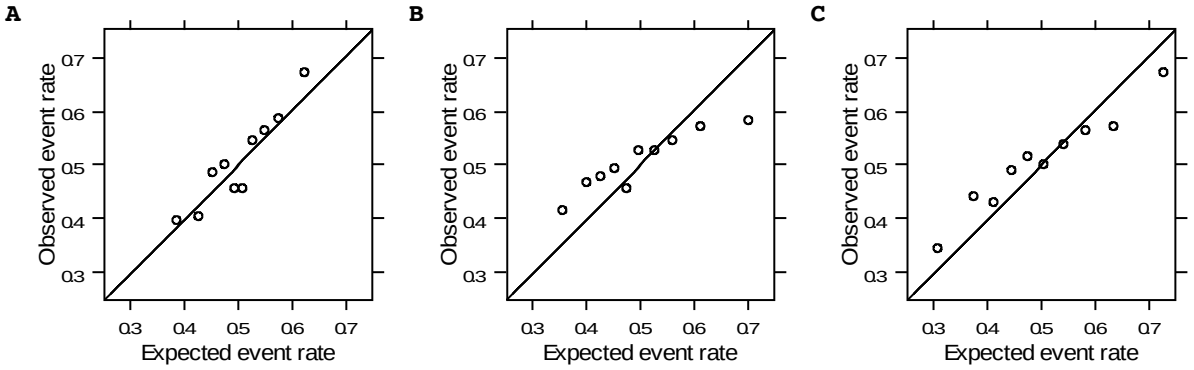
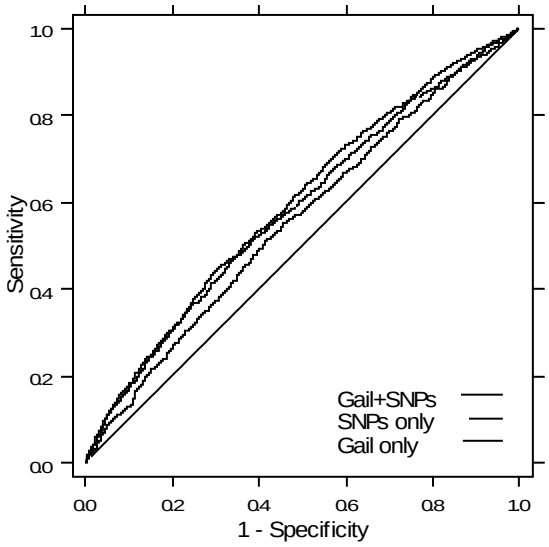


Figure 3.



Supplemental Methods and Data

Supplemental Methods

Genotyping

The samples were genotyped on custom oligonucleotide arrays across 9039 SNPs selected from previous genome-wide association studies, including 6 of the 7 replicated breast cancer association loci. Pair-wise identity-by-state analysis identified three apparent sibling pairs, and we excluded one member of each pair from analyses. The average concordance across duplicate samples included for quality control was 99.8%, and the breast cancer loci all had call rates above 99%. We used principal components analysis to model population structure, and results were generally consistent with self-reported ethnicity.

Separately, samples with sufficient available DNA were genotyped across all 7 breast cancer loci by Sequenom on the MassArray platform, along with 16 additional SNPs for quality control that had also been genotyped on the arrays. We designed two assays in opposing orientations for each of the breast cancer SNPs, and these were carried out in separate multiplexes. We used whole genome amplification (WGA) for roughly 70% of the samples and observed a reduction in genotyping quality for the WGA samples as compared to the genomic DNA samples. While most WGA samples had satisfactory performance, a subset showed a combination of elevated missing data rates, reduced heterozygosity, and inconsistencies with the array-based genotype data. As a result, we excluded Sequenom data for any sample that had

more than one inconsistency with a corresponding array-based genotype. This led to exclusion of Sequenom genotype data for 121 samples, which were almost equally distributed across cases and controls.

We scored consensus genotypes for rs13387042 by combining calls from the two Sequenom assays, scoring conflicting genotypes as missing. Concordance across the two assays was 97%, and the resulting call rate for the consensus genotypes was 96.8% across samples that had not been excluded for lack of DNA or poor data quality. While these assays performed relatively poorly, the consensus of the two assays should still be very accurate, and missing consensus calls were not differentially distributed across cases and controls ($P = 0.27$, χ^2 test).

The Composite SNP Risk Score

Based on a log-additive risk model, the three genotypes AA, AB, and BB for a single SNP have relative risk values of 1, OR, and OR^2, under a rare disease model where OR is the odds ratio for the high risk B allele. If the B allele has frequency p, then these genotypes have population frequencies of $(1-p)^2$, $2p(1-p)$ and p^2, assuming Hardy Weinberg equilibrium. We scaled the genotype relative risk values for each SNP so that based on these frequencies, the average relative risk in the population is 1.

We considered two approaches for combining SNP risk with Gail risk estimates. The Gail model consists of a relative risk estimate based on clinical risk factors, projected to absolute risk

based on demographic tables of breast cancer incidence and competing mortality rates. The most accurate method for incorporating SNP risk is to use SNP risk to adjust the Gail relative risk, and then apply the Gail method to derive absolute risk. A more convenient approach in some cases may be to use the Gail model unchanged and then multiply the final absolute risk by the SNP relative risk; this is an appropriate approximation when the absolute disease risk is small. The two approaches give similar results when absolute risk is estimated over short time intervals (such as 5 years) and differences become apparent only at high risk levels (Supplemental Figure 1). For this study, we used the second estimation approach. Thus, the formula for our combined SNP×Gail absolute risk score is:

$$SNP{\times}Gail = Gail \quad SNP_1 \quad SNP_2 \quad SNP_3 \quad SNP_4 \quad SNP_5 \quad SNP_6 \quad SNP_7$$

where *Gail* is the Gail absolute risk score, and $SNP_{1..7}$ are relative risk scores for the individual SNPs, each scaled to have a population average of 1.

Supplemental Data

Genotyping Performance Summary

Supplemental Table 1 summarizes genotyping results for the 7 BCa associated SNPs. We excluded from analyses 27 samples that had more than 2 missing genotypes out of these 8 SNPs; for the remaining samples included in our analyses, 90% had complete data for the 8 SNPs. The lower call rate for rs13387042 is primarily a result of the samples that could not be

genotyped due to insufficient DNA. There were no apparent deviations from Hardy-Weinberg equilibrium.

Estrogen Receptor Focused Analysis

A previous study had shown that the Gail model is more effective at predicting estrogen receptor positive (ER+) than ER negative (ER-) tumors in the WHI cohort (S1), and most of the individual SNPs have shown stronger associations with ER+ than ER- tumors (S2). We used logistic regression to measure association for the risk scores by tumor subtype (Supplemental Table 3). Results for ER+ tumors were similar to results for all invasive breast cancers. Gail risk and the combined risk score were not predictive for ER- tumors; SNP risk by itself still had evidence for association ($P = 0.04$), but the effect size was poorly defined due to the smaller sample size. We used bootstrap resampling to assess significance of differences in AUC for risk scores as a function of tumor receptor status (Supplemental Table 4). Gail and combined risk scores had substantially larger AUC for ER+ than ER- tumors (empirical $P < 0.001$ in both cases). For SNP risk, the difference in AUC for ER+ versus ER- tumors was borderline significant (95% CI: 0.01 – 0.10, two-sided $P = 0.02$). AUC for ER+ tumors was also significantly larger than for all invasive cancers for both Gail score (empirical $P = 0.02$) and the combined score (empirical $P = 0.02$), though these differences were small.

We investigated using ER-specific odds ratios to form separate ER-positive and ER-negative versions of the SNP risk score. We compared performance of these scores to the

general SNP risk score by AUC. Bootstrap confidence intervals for the difference in AUC crossed zero for both receptor status subtypes (Supplemental Table 5). While we were unable to conclusively demonstrate improved classification with the available sample sizes, prediction of ER- cancers appeared most likely to improve (one-sided $P = 0.05$).

Supplemental References

S1. Chlebowski RT, Anderson GL, Lane DS, *et al*. Predicting risk of breast cancer in postmenopausal women by hormone receptor status. *J Natl Cancer Inst* 2007;99:1695-705.

S2. Garcia-Closas M, Chanock S. Genetic susceptibility loci for breast cancer by estrogen receptor status. *Clin Cancer Res* 2008;14:8000-9.

Supplemental Table 1. Genotyping performance of the breast cancer loci

dbSNP rsID	Platform	Call rate	Freq*	P_{HWE} †
rs2981582	Array	1.000	0.406	0.29
rs3803662	Array	1.000	0.288	0.85
rs889312	Array	0.996	0.285	0.64
rs13387042	Sequenom	0.910	0.512	0.58
rs13281615	Array	1.000	0.420	0.43
rs4415084	Array	0.997	0.401	0.07
rs3817198	Array	1.000	0.322	0.60

* Frequency of the previously-reported high risk allele.

† *P* value for Hardy-Weinberg equilibrium, from a likelihood ratio test.

Supplemental Table 2. Expected and observed counts of cases and controls for deciles of risk scores.

	Expected		Observed	
Risk Decile	Control	Case	Control	Case
SNP risk				
[0.454,0.687]	209.3	132.7	207	135
(0.687,0.772]	191.4	144.6	201	135
(0.772,0.848]	170.2	141.8	161	151
(0.848,0.921]	184.8	168.2	177	176
(0.921,0.991]	163.6	161.4	178	147
(0.991,1.06]	151.8	160.2	170	142
(1.06,1.14]	157.1	177.9	153	182
(1.14,1.25]	146.1	179.9	143	183
(1.25,1.43]	140.7	193.3	139	195
(1.43,2.65]	121.0	204.0	107	218
Gail risk				
[0.0066,0.011]	212.6	118.4	194	137
(0.011,0.012]	195.4	132.6	175	153
(0.012,0.014]	190.7	144.3	176	159
(0.014,0.015]	178.7	149.3	167	161
(0.015,0.016]	174.9	160.1	183	152
(0.016,0.018]	161.7	163.3	155	170
(0.018,0.02]	154.9	174.1	156	173
(0.02,0.024]	143.8	185.2	150	179
(0.024,0.032]	126.8	203.2	142	188
(0.032,0.12]	96.7	233.3	138	192
SNP×Gail risk				
[0.0045,0.0092]	226.9	103.1	217	113
(0.0092,0.011]	204.9	125.1	185	145
(0.011,0.013]	192.5	137.5	189	141
(0.013,0.014]	182.5	147.5	169	161
(0.014,0.016]	172.5	157.5	161	169
(0.016,0.018]	161.9	168.1	166	164
(0.018,0.022]	150.3	179.7	153	177
(0.022,0.026]	136.4	193.6	145	185
(0.026,0.033]	119.6	210.4	142	188
(0.033,0.23]	88.4	241.6	109	221

Supplemental Table 3. Reclassification table for SNP×Gail risk versus Gail risk in women with a previous breast biopsy.

	SNP risk × Gail 5-Year risk			
Gail 5-Year risk	< 1.5%	1.5% - 2.0%	> 2.0%	Total
< 1.5%				
Women	71	23	5	99
Events	39	15	4	58
Nonevents	32	8	1	41
Proportion	0.549	0.652	0.800	0.586
1.5% - 2.0%				
Women	58	65	65	188
Events	29	38	48	115
Nonevents	29	27	17	73
Proportion	0.500	0.585	0.738	0.612
> 2.0%				
Women	18	46	320	384
Events	7	20	200	227
Nonevents	11	26	120	157
Proportion	0.389	0.435	0.625	0.591
Total				
Women	147	134	390	671
Events	75	73	252	400
Nonevents	72	61	138	271
Proportion	0.510	0.545	0.646	0.596

Supplemental Table 4. Logistic regression tests of association with estrogen receptor subtypes.

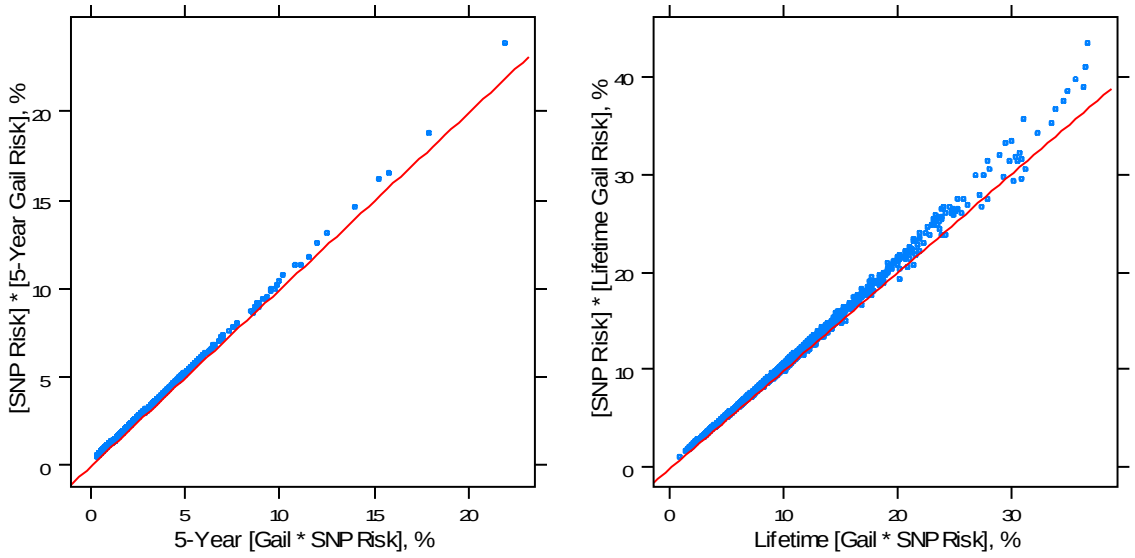
	ER-positive tumors		ER-negative tumors	
Predictor	(95% CI)	*P*	(95% CI)	*P*
log(Gail 5-year risk)	0.55 (0.37 – 0.72)	1.1×10^{-9}	-0.03 (-0.37 – 0.32)	0.89
log(SNP risk)	1.20 (0.92 – 1.47)	1.7×10^{-18}	0.56 (0.03 – 1.09)	0.04
log(SNP × Gail risk)	0.72 (0.57 – 0.87)	2.4×10^{-22}	0.14 (-0.14 – 0.43)	0.32

Supplemental Table 5. AUC and bootstrap 95% confidence intervals for risk scores versus tumor receptor status.

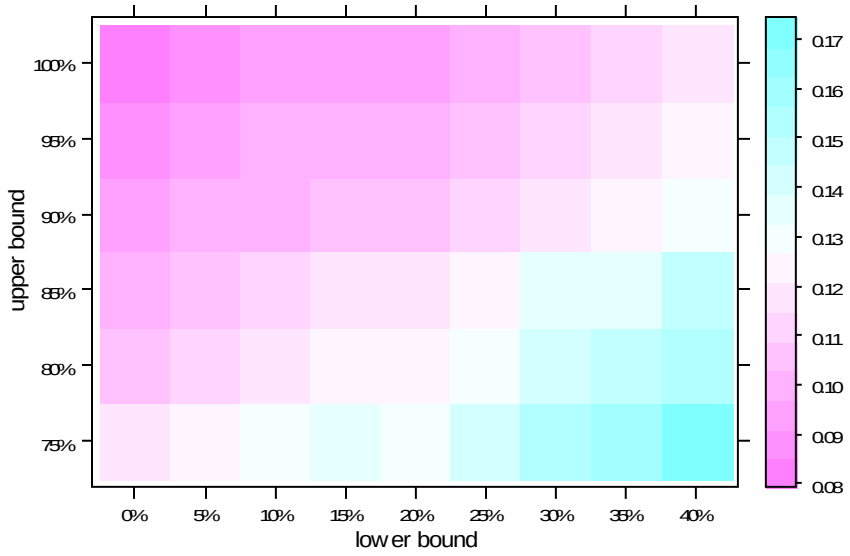
Predictor	All tumors	ER-positive	ER-negative
log(Gail 5-year risk)	0.557 (0.537 – 0.575)	0.568 (0.547 – 0.587)	0.486 (0.445 – 0.532)
log(SNP risk)	0.587 (0.567 – 0.607)	0.593 (0.572 – 0.614)	0.541 (0.496 – 0.583)
log(SNP × Gail risk)	0.594 (0.575 – 0.612)	0.605 (0.583 – 0.625)	0.521 (0.478 – 0.567)

Supplemental Table 6. AUC and bootstrap 95% confidence intervals for receptor-specific risk scores.

Tumor type	General score	ER-specific score	Difference in AUC
ER-positive	0.593 (0.572 – 0.614)	0.591 (0.570 – 0.613)	-0.002 (-0.005 – 0.001)
ER-negative	0.541 (0.496 – 0.583)	0.564 (0.522 – 0.602)	0.022 (-0.005 – 0.050)



Supplemental Figure 1. Comparison of approaches for combining Gail and single nucleotide polymorphism (SNP) risk. The X axes show risk estimates obtained by multiplying Gail relative risk by SNP relative risk, and then converting to absolute risk. The Y axes show corresponding estimates obtained by multiplying Gail absolute risk by SNP relative risk. The results are generally similar, though the second method tends to slightly overestimate very high risk values. For the 5-year risk estimates, $r^2 = 0.9995$, and for lifetime risk, $r^2 = 0.9972$.



Supplemental Figure 2. Net reclassification improvement (NRI) excluding the upper and lower tails of the distribution of Gail absolute risk. The lower and upper bounds represent quantiles of Gail risk in the full cohort. The upper left corner represents NRI for the entire cohort, and the lower right represents the most restrictive subset with a lower bound near 1.5% and an upper bound near 2.0%.