Genetic Variants in Metabolic Signaling Pathways and Their Interaction with Lifestyle Factors on Breast Cancer Risk: A Random Survival Forest Analysis

Su Yon Jung[1], Jeanette C. Papp[2], Eric M. Sobel[2], and Zuo-Feng Zhang[3]

Abstract

Genetic variants in the insulin-like growth factor-I (IGF-I)/insulin resistance axis may interact with lifestyle factors, influencing postmenopausal breast cancer risk, but these interrelated pathways are not fully understood. In this study, we examined 54 single-nucleotide polymorphisms (SNP) in genes related to IGF-I/insulin phenotypes and signaling pathways and lifestyle factors in relation to postmenopausal breast cancer, using data from 6,567 postmenopausal women in the Women's Health Initiative Harmonized and Imputed Genome-Wide Association Studies. We used a machine-learning method, two-stage random survival forest analysis. We identified three genetic variants (*AKT1* rs2494740, *AKT1* rs2494744, and *AKT1*

rs2498789) and two lifestyle factors [body mass index (BMI) and dietary alcohol intake] as the top five most influential predictors for breast cancer risk. The combination of the three SNPs, BMI, and alcohol consumption (≥ 1 g/day) significantly increased the risk of breast cancer in a gene and lifestyle dose-dependent manner. Our findings provide insight into gene–lifestyle interactions and will enable researchers to focus on individuals with risk genotypes to promote intervention strategies. These data also suggest potential genetic targets in future intervention/clinical trials for cancer prevention in order to reduce the risk for breast cancer in postmenopausal women. *Cancer Prev Res; 11(1); 44–51. ©2017 AACR.*

Introduction

Breast cancer is the most commonly occurring cancer in women and the second most common cause of cancer-related deaths in the United States (1, 2). Incidence and death rates for breast cancer increase with age. Approximately 79% of new cases and 88% of cancer deaths occur in women age 50 years and older (2, 3). The insulin-like growth factor-I (IGF-I)/insulin resistance (IR) axis demonstrates strong associations with breast cancer (4–6). Total and/or free bioavailable IGF-I proteins are associated with higher risk of breast cancer and worse cancer survival in premenopausal and postmenopausal women (7–9). In postmenopausal women, homeostatic model assessment–insulin resistance (HOMA-IR), as a proxy measure of IR, reflecting compensatory high blood levels of insulin and glucose, is positively associated with breast cancer

(10). Hyperinsulinemia alone has been associated with a 2-fold increase in breast cancer risk in postmenopausal women (11, 12).

High IGF-I levels and IR (characterized by hyperinsulinemia and hyperglycemia) activate the IGF/insulin receptors, which are overexpressed in breast cancer cells. This overexpression results in the enhanced anabolic state necessary for cell proliferation, differentiation, and antiapoptosis via deregulating or overactivating multiple downstream signaling pathways, including the PI3K/protein kinase B (Akt) and MAPK pathways (13–15). Thus, high IGF-I levels and IR contribute to overexpression of relevant receptors and to hyperactive and abnormal multiple cell-signaling pathways, and therefore may be associated with carcinogenesis.

Considering the associations of the IGF-I/IR axis with breast cancer risk, the genetic variants that may influence circulating levels of IGF-I and insulin are possibly associated with breast cancer risk. In addition, the IGF-I/insulin signaling pathways' genetic alterations lead to altered gene expression and protein function and are plausibly associated with increased risk of breast cancer. However, population-based studies to examine these genetic variants and breast cancer relationships have been limited and yield inconsistent findings (16–23).

Behavioral factors may interact with genetic factors and jointly influence breast cancer susceptibility. In postmenopausal women, obesity is associated with increased risk of breast cancer, which could be mediated via the IGF-I/IR axis (4). Also, an unhealthy, unbalanced diet could be a potential risk factor for breast cancer. In particular, alcohol, even at a few drinks a week, could increase breast cancer risk (24, 25).

Gene–behavior interaction is a critical area in cancer genetic epidemiology and has been studied with various statistical methods. In this retrospective study among non–Hispanic white

[1]Translational Sciences Section, Jonsson Comprehensive Cancer Center, School of Nursing, University of California, Los Angeles, Los Angeles, California. [2]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California. [3]Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, California.

Corresponding Author: Su Yon Jung, University of California Los Angeles, 700 Tiverton Avenue, 3-264 Factor Building, Los Angeles, CA 90095. Phone: 310 825-2840; Fax: 310 267-0413; E-mail: sjung@sonnet.ucla.edu

AACR

postmenopausal women, we evaluated 54 single-nucleotide poly-morphisms (SNP) in genes related to the IGF-I/insulin pheno-types and signaling pathways and selected 17 demographic and lifestyle factors. We evaluated the genetic variants and lifestyle factors by ranking them according to their predictive value and accuracy for breast cancer. We then examined the effect of inter-action between the most influential genetic variants and lifestyle factors on predicting breast cancer risk. We used a machine learning method, two-stage random survival forest (RSF) analysis. The recently developed RSF tool is a nonparametric tree-based ensemble learning method and accounts for the nonlinear effects of variables that may not be handled in a regression model (26, 27). This allows for high-order interactions among variables and has yielded accurate predictions (26). Thus, this method may provide a way to resolve the conflicting findings in previous studies of genes and behaviors. By applying the two-stage RSF approach, we tested the hypothesis that the most dominant genetic and behavioral factors identified through the RSF analysis interact reciprocally to predict breast cancer risk. We further evaluated a gene and behavior dose-response relationship and estimated the combined effect of those variables on breast cancer risk.

Materials and Methods

Study population

This study included data from 6,567 participants enrolled in the Women's Health Initiative (WHI) Harmonized and Imput-ed Genome-Wide Association Studies (GWAS) data, which contributes a joint imputation and harmonization effort for GWAS within the WHI Clinical Trials and Observational Stud-ies. Details of the studies' rationale and design have been described elsewhere (28, 29). Briefly, WHI study participants were recruited from 40 clinical centers nationwide from Octo-ber 1, 1993, to December 31, 1998. Eligible women were 50 to 79 years old, postmenopausal, expected to live near the clinical centers for at least 3 years after enrollment, and able to provide written consent. For our study, we initially included 10,703 of those women who reported their race or ethnicity as non–Hispanic white (Supplementary Fig. S1). Of those, we excluded 469 women who had been followed up for less than 1 year or had been diagnosed with any cancer at enrollment. We also excluded women ($n = 1,793$) who had diabetes mellitus at enrollment or later. We excluded another 1,101 women whose SNP data indicated they were duplicated or related to others in the dataset. Of the 7,340 women remaining, we finally excluded 773 women for whom the information on covariates was unavailable, resulting in a total of 6,567 women (90% of the eligible 7,340). Of these, 352 developed breast cancer after enrollment. The participants had been followed up through August 29, 2014 (a median follow-up period of 16 years). This study was approved by the institutional review boards of each participating clinical center of the WHI and the University of California, Los Angeles.

Data collection and cancer outcome variables

Data had been collected using standardized written protocols with periodic quality assurance (QA) performed by the WHI coordinating center. At baseline, participants completed self-administered questionnaires regarding demographic factors (age, family income, and family history of breast cancer),

lifestyle factors [depressive symptom, smoking status, physical activity, diet (dietary alcohol in grams per day and percent calories from fat and from saturated fatty acids per day)], and reproductive histories [exogenous estrogen (E) use (never vs. duration of E only and E + Progestin (P) use), history of hysterectomy or oophorectomy, ages at menopause and men-arche, and number of pregnancies]. Anthropometric measure-ments such as height, weight, and waist and hip circumferences were measured at baseline by trained staff. The above variables were initially selected for this study on the basis of a literature review for their associations with breast cancer. Multicollinear-ity testing and univariate and stepwise regression analyses determined the final set to be analyzed.

Cancer outcomes were determined using a centralized review of medical charts, and cancer cases were coded according to the National Cancer Institute's Surveillance, Epidemiology, and End-Results guidelines (30). The outcome variables were breast cancer and the time to development of breast cancer. The time from enrollment to breast cancer development, censoring, or study endpoint was estimated as the number of days and then converted into years.

Genotyping

The WHI Harmonized and Imputed GWAS is a combination of six substudies (Hip Fracture GWAS, SHARe, GARNET, WHIMS, GECCO, and MOPMAP) within the WHI study. Genotyping included alignment ("flipping") to the same reference panel and imputation via the 1,000 Genomes reference panels. SNPs for harmonization were checked for pairwise concordance and for identity by descent in Plink to identify relatedness among all samples in the substudies. Initial QA was implemented according to a standardized protocol, with 90% R^2 imputation quality scores, a missing call rate of <2%, and a Hardy–Weinberg equi-librium of $P \geq 10^{-4}$. Fifty-four SNPs in 9 genes (Supplementary Table S1) were chosen based on the biological significance of their gene products, or whether epidemiologic and/or experimental data support an association between the gene and the levels of IGF and insulin, or between the gene and risk of cancer (13, 16–23, 31–33). The allele frequencies of these SNPs in our population were consistent with the frequencies in a European population (ref. 34; http://browser.1000genomes.org).

Statistical analysis

Differences in baseline characteristics and allele frequencies by breast cancer status were evaluated by using unpaired two-sample t tests for continuous variables and χ^2 tests for categorical vari-ables. If continuous variables were skewed or had outliers, Wil-coxon rank-sum test was used. The Cox proportional hazards regression model was conducted to obtain hazard ratios (HR) and 95% confidence intervals (CI) for IGF-I/insulin–related SNPs (as a categorical variable of an additive model and major-allele dom-inant model) and for the combined effect of the SNPs and lifestyle factors in predicting breast cancer.

The RSF analysis involves obtaining bootstrap samples from the original cohort and growing a tree for each bootstrapped sample, based on a splitting rule applied to a tree node to maximize survival differences across daughter nodes. The process is repeated numerous times (number of trees = 5,000 in this study) so that a forest of trees is created (35, 36). An ensemble cumulative hazard estimate for each individual was calculated from each tree and averaged over all trees, yielding a predicted

Table 1. Characteristics of participants, stratified by breast cancer

Characteristic	Controls (n = 6,215) n (%)	Breast cancer cases (n = 352) n (%)
Age in years, median (range)	67 (50–81)	68 (50–79)
Family income		
<$35,000	2,941 (48.3)	155 (44.8)
≥$35,000	3,145 (51.7)	191 (55.2)
Family history of breast cancer		
No	5,200 (83.7)	275 (78.1)[a]
Yes	1,015 (16.3)	77 (21.9)
Depressive symptom[b], median (range)	0.002 (0.000–0.919)	0.002 (0.000–0.880)
Dietary alcohol per day in g, median (range)	1.02 (0.00–153.60)	1.57 (0.00–106.70)[a]
% calories from fat, median (range)	33.58 (7.81–65.54)	34.15 (11.71–60.35)
% calories from SFA, median (range)	11.21 (2.22–30.80)	11.76 (3.86–20.10)
METs·hour/week[c]	2.25 (0.00–142.30)	0.63 (0.00–54.33)[a]
Smoking now		
No	5,943 (95.6)	331 (94.0)
Yes	272 (4.4)	21 (6.0)
BMI in kg/m^2, median (range)	27.01 (15.42–58.49)	28.30 (18.31–47.67)[a]
Waist-to-hip ratio, median (range)	0.81 (0.44–1.26)	0.82 (0.64–1.07)
Age at menarche in years, median (range)	13 (≤9–≥17)	12 (≤9–≥17)[a]
Age at menopause in years, median (range)	50 (23–71)	50 (21–63)
Number of pregnancies, median (range)	3 (0–8)	3 (0–8)
History of hysterectomy or oophorectomy		
No	3,831 (61.6)	241 (68.5)[a]
Yes	2,384 (38.4)	111 (31.5)
Exogenous estrogen use (E only use)		
Never use	4,427 (71.2)	275 (78.1)[a]
<5 years	948 (15.3)	32 (9.1)
5 to <10 years	303 (4.9)	12 (3.4)
10 to <15 years	213 (3.4)	15 (4.3)
15+ years	324 (5.2)	18 (5.1)
Exogenous estrogen use (E + P use)		
Never use	5,278 (84.9)	276 (78.4)[a]
<5 years	582 (9.4)	44 (12.5)
5 to <10 years	188 (3.0)	16 (4.5)
10 to <15 years	99 (1.6)	13 (3.7)
15+ years	68 (1.1)	3 (0.9)

Abbreviations: BMI, body mass index; E, estrogen; E + P, estrogen + progestin; MET, metabolic equivalent; SFA, saturated fatty acids
[a]$P < 0.05$, χ^2 or Wilcoxon rank-sum test.
[b]Depression scales were estimated using a short form of the Center for Epidemiologic Studies Depression Scale.
[c]Physical activity was estimated from recreational physical activity combining walking and mild, moderate, and strenuous physical activity; each activity was assigned a MET value corresponding to intensity, and the total MET·hours/week was calculated by multiplying the MET level for the activity by the hours exercised per week and summing the values for all activities (46).

cumulative incidence rate of breast cancer. The prediction algorithm was applied to the out-of-bag (OOB) data (37% of the original data not used for bootstrapping) to calculate the OOB concordant index (c-index), a measure of prediction performance, which is conceptually similar to the area under the ROC curve (AUC; refs. 35, 37). The importance of each variable was determined by two predicted values: (i) minimal depth, where variables with a small minimal depth split the tree close to the root and are considered highly predictive and (ii) variable importance (VIMP), calculated as the difference between the OOB c-indexes from the original OOB data and from the permuted OOB data, where variables with larger VIMP are the more predictive (26, 38).

We used a two-stage RSF approach. In the first stage, we performed an RSF on each SNP and each lifestyle factor individually (Supplementary Tables S2 and S3; Supplementary Figs. S2 and S3); only those SNPs with significantly low minimal depth and high VIMP scores were selected for the second stage. During stage two, we performed another RSF using all lifestyle factors but only the SNPs selected during stage one. All lifestyle factors were used in the second stage because their rank did not change to reduce noise in both stages. This method allows us to eliminate the SNPs that may not have effects on predicting breast cancer, which will result in more statistical power with the correct type I error than the original RF-based analysis (36). A P value < 0.05 was considered statistically significant. R version 3.3.2 with survival, randomForestSRC, ggRandomForests, and gamlss packages were used.

Results

Participants' baseline characteristics by breast cancer status are presented in Table 1. Women with breast cancer were more likely to have a family history of breast cancer, consume more dietary alcohol per day, be inactive, and have greater body mass index (BMI). In addition, women with breast cancer tended to have undergone earlier menarche and were less likely to have a history of hysterectomy or oophorectomy. Finally, women with breast cancer had a lower rate of exogenous E-only use and shorter

Table 2. Prediction of variable using the RSF model

Variable[a]	Predictive value[b]	VIMP
AKT1 rs2494740	2.7162	0.0066
BMI	2.8566	0.0011
AKT1 rs2494744	2.9388	0.0090
AKT1 rs2498789	3.1114	0.0075
Dietary alcohol per day	3.1386	0.0016
Age at menopause	3.2740	0.0002
Depressive symptom	3.4122	0.0004
AKT1 rs1130214	3.6116	0.0030
Waist-to-hip ratio	3.8390	−0.0005
IRS1 rs1801278	4.0562	0.0066
Percent calories from SFA per day	4.0592	0.0011
Percent calories from fat per day	4.1002	−0.0007
Physical activity	4.8408	0.0003
Age	4.9082	0.0002
Age at menarche	4.9662	0.0001
Family income	5.4110	0.0000
Number of pregnancies	5.8770	−0.0001
E + P use	7.8110	0.0008
Family history of breast cancer	8.2020	0.0006
History of either hysterectomy or oophorectomy	10.1706	0.0002
Smoking status	10.2046	0.0000
E only use	11.2790	−0.0001

Abbreviations: BMI, body mass index; E, estrogen; P, progestin; SFA, saturated fatty acids; VIMP, variable of importance.
[a]Variables are ordered by predictive value.
[b]Predictive value of variable was assessed via minimal depth method in the nested RSF models. A lower value is likely to affect greatly prediction.

duration of use, but had a higher rate of E + P use and longer duration of use.

The most influential variables for breast cancer risk identified via minimal depth and VIMP

In the two-stage RSF analysis, we used two predicted measures to identify the most influential variables (i.e., having the highest predictive value and least prediction error). After selecting the most influential SNPs at the first stage (Supplementary Table S3 and Supplementary Fig. S3), we then performed the second RSF on the five selected SNPs and all 17 lifestyle factors to predict breast cancer risk. The minimal depth and VIMP measures use different criteria, so we expected the variable ranking to be somewhat different. We thus estimated those values in Table 2 and compared the two measures using Fig. 1A. In the plot, variables were sorted via the minimal depth's rank in the _y_-axis, and points are colored and shaped by the sign of VIMP. The red dashed line indicates where the two measures were in agreement: The further the points were from the line, the more the discrepancy between measures. In this figure, both minimal depth and VIMP indicate the following three genetic variants and two lifestyle factors are strong predictive markers of breast cancer risk: _AKT1_ rs2494740, _AKT1_ rs2494744, and _AKT1_ rs2498789, BMI, and dietary alcohol intake per day.

The OOB c-index (Fig. 1B) for the nested RSF model orders variables according to their predictive value assessed via the minimal depth method. Results indicated that the above top five variables (three SNPs and two lifestyle factors) improved the overall OOB c-index and thus had complementary predictive value, while others did not add to a significant improvement of the prediction accuracy.

Cumulative incidence rate of breast cancer for the most influential variables and their cumulative effects on breast cancer risk

To account for the nonlinear effects of variables on cancer risk, the predicted cumulative incidence rate of breast cancer for the top five variables were estimated based on the RSF model (Fig. 2A–E). The genotype of each SNP was analyzed as a continuous variable.

The cumulative effects of the three SNPs and two lifestyle factors were further calculated and shown in Table 3. Based on the results in Fig. 2A–C, the genotypes of _AKT1_ rs2494740 AA, _AKT1_ rs2494744 AA, and _AKT1_ rs2498789 GG were determined as risk genotypes and analyzed as categorized variables in Table 3. In Fig. 2E, the BMI had a U-shaped risk for breast cancer, diverging from around 30 kg/m^2; we thus stratified by BMI using 30 kg/m^2 as a cutoff value and obtained the joint effect of BMI with the three SNPs and alcohol intake on cancer risk.

In an individual SNP analysis (Supplementary Table S4) with additive and major-allele dominant models, no significant associations were found; however, the combination of the SNPs in Table 3 provided different results. Compared with nonobese women with null risk genotypes, obese women carrying one or more risk genotypes had higher risk of breast cancer (HR 5.65; 95% CI, 2.08–15.36). Consistently, obese women who consumed dietary alcohol ≥1 g/day had higher breast cancer risk than nonobese women who consumed alcohol <1 g/day. Furthermore, compared with women with null risk genotypes and alcohol consumption <1 g per day, those with both factors had higher risk of breast cancer, suggesting the cumulative effect of genetic and lifestyle factors. When stratified by BMI, obese women with one and both factors of risk genotypes and alcohol intake (≥1 g/day) had a 2-fold and 7-fold increased breast cancer risk, respectively, compared with nonobese women with null risk genotypes and alcohol consumption <1 g/day. These results indicate a gene and lifestyle dose–response relationship and significant joint effect of BMI with the SNPs and alcohol consumption on cancer risk.

Discussion

Using the two-stage RSF approach, we identified three genetic variants (_AKT1_ rs2494740, _AKT1_ rs2494744, and _AKT1_ rs2498789) and two lifestyle factors (BMI and dietary alcohol intake) as the top five most influential predictors for breast cancer risk in this dataset of postmenopausal women. We further examined interaction effects of those factors on cancer risk. In the individual SNP analysis, no significant association was observed, but the combination of the three SNPs in addition to BMI and alcohol intake significantly increased the risk of breast cancer.

The PI3K/Akt pathway leads to metabolic activity, including glucose uptake and decreased apoptosis and is a main signaling cascade in controlling the cellular process promoting carcinogenesis (39). Two members of the AKt family, AKt1 and AKt2, are important signaling molecules related to a diabetic phenotype such as IR; in addition, at the genomic level, each is amplified in various cancers including breast cancer (40, 41). The _AKT1/2_ genes are thus key components of this pathway, but studies of the association of their genetic variants with breast cancer have been limited (13, 32). In our study of 10 SNPs in the _AKT1/2_ genes, three SNPs were identified as the top three most influential
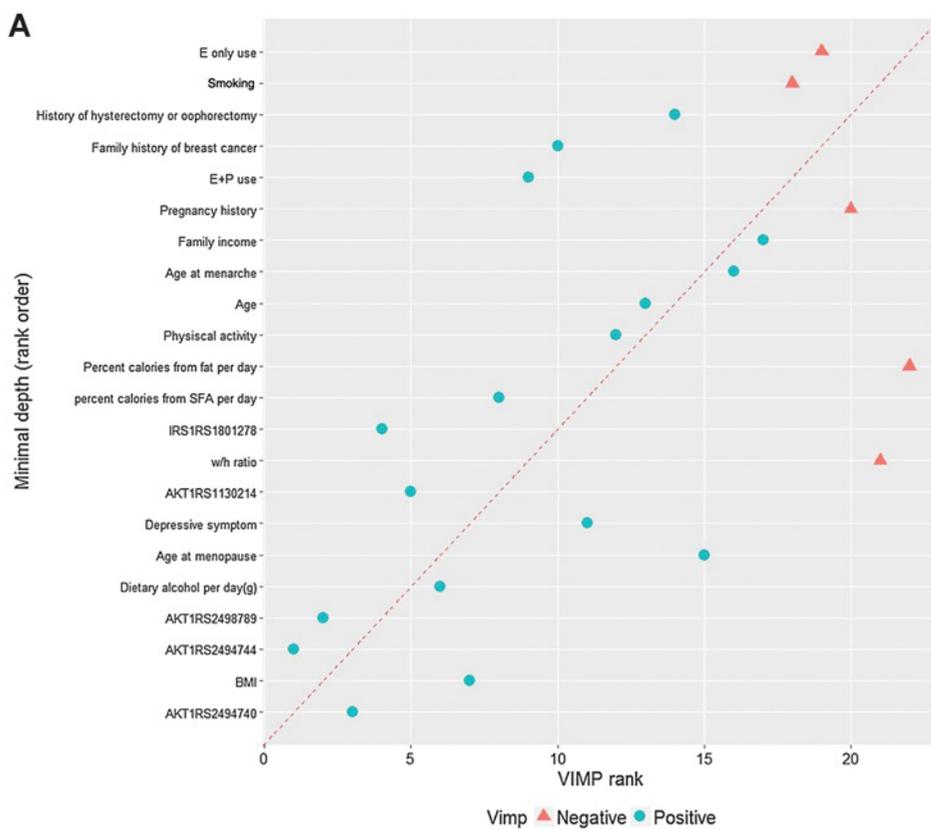
Jung et al.

Figure 1.
Predictive value of variable. **A,** Comparing minimal depth and VIMP rankings. (BMI, body mass index; E, exogenous estrogen; P, progestin; SFA, saturated fatty acids; VIMP, variable of importance; w/h ratio, waist-to-hip ratio). **B,** OOB concordance index. [Improvement in OOB concordant index (c-index) was observed when the top five variables (●) were added to the model, whereas other variables (○) did not further improve the accuracy of prediction.]
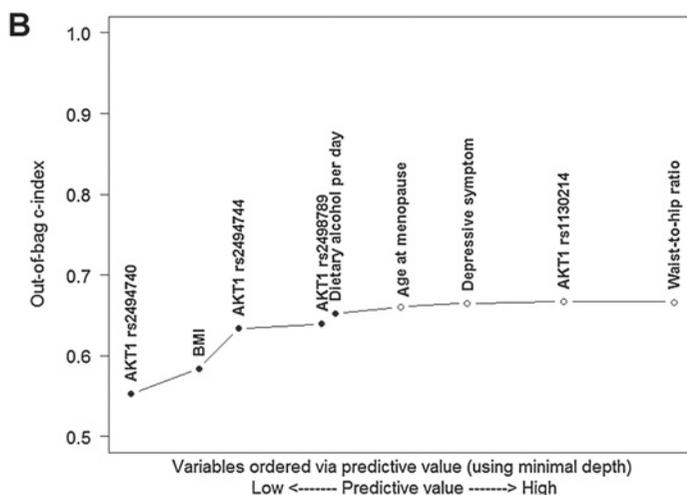
genetic factors. However, studies of the functional biology of these SNPs have been limited, warranting further study.

Alcohol use is causally associated with several cancers, and even a low dose (≤1 drink/day) increases the risk of breast cancer (24, 25, 42). Consistently, we found that women who consumed ≥1 g/day of alcohol had higher risk of breast cancer; furthermore, in obese women, the low-dose consumption level (≥1 g/day) caused a 2-fold increased risk for breast cancer, compared with nonobese women consuming <1 g/day of alcohol.

Combined with alcohol intake of ≥1 g/day, the effects of the three SNPs significantly strengthened, suggesting the cumulative interacting effect of those genetic and lifestyle factors on breast cancer risk. In addition, in obese women, those factors were associated with breast cancer risk in a gene and lifestyle dose-dependent manner, indicating a joint effect of BMI with those factors on breast cancer risk.

The self-reporting of the dietary alcohol intake, smoking, and physical activity data limits study conclusions regarding these
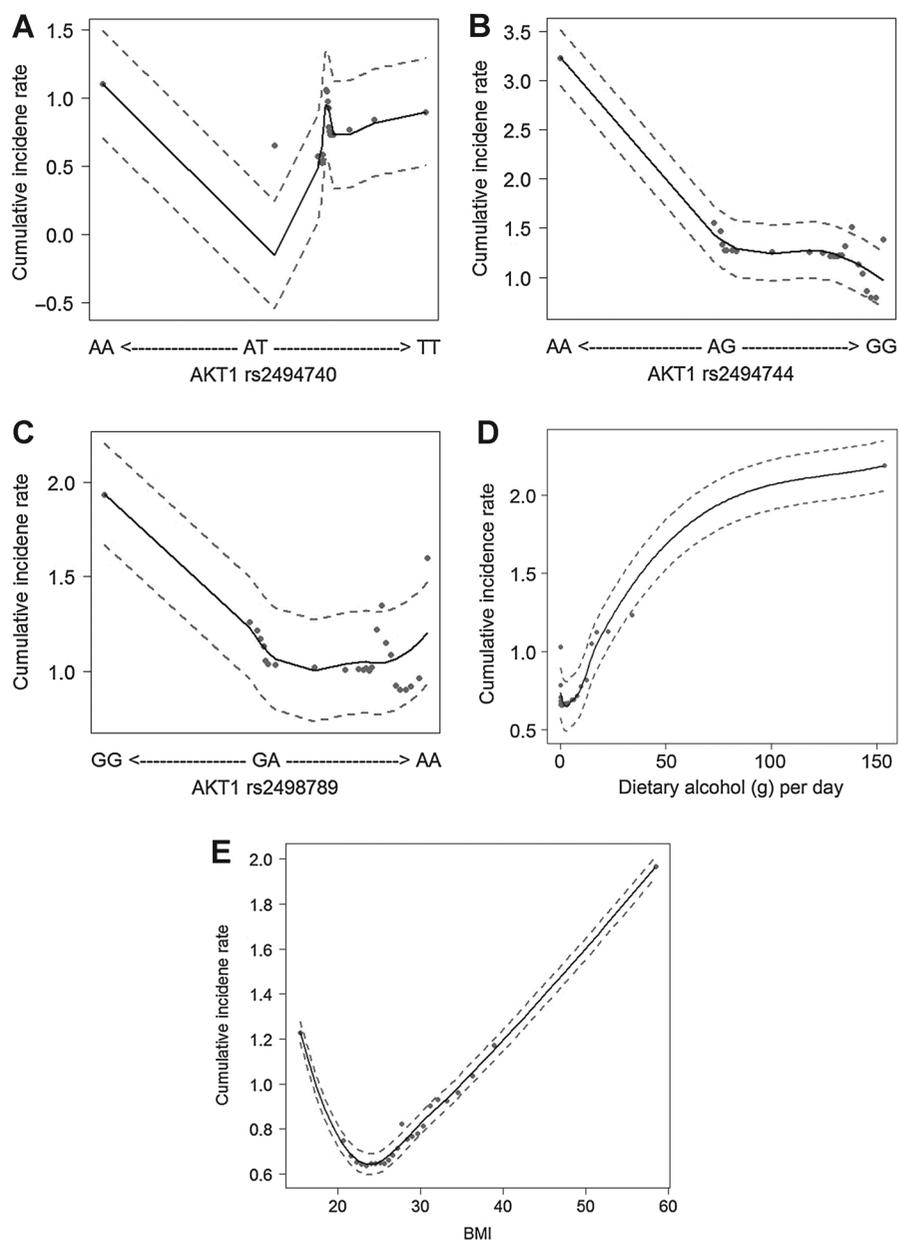
Figure 2.
Cumulative breast cancer incidence rate for the five most influential variables based on an RSF analysis. Dashed gray lines indicate 95% CI.

variables due to the likely prevalence of underreporting of alcohol intake and smoking and overreporting of physical activity, especially in obese women. Our study population included data from non–Hispanic white postmenopausal women only, so the generalizability of our findings to other populations is limited. We acknowledge that the statistical power for detecting gene–environment interaction was relatively low in this study; we conducted a two-stage RSF analysis to have more statistical power with the correct type I error than the original RF-based analysis. Despite these limitations, the potential impact of our findings clearly warrants further study. We used a two-stage RSF method to identify the most predictive variables for breast cancer risk. The RSF provides a robust way to handle high-level interactions in variables and allows for accurate prediction. In several research areas, including molec-

ular genetic epidemiology, this method has outperformed the traditional models by accounting for the nonlinear effects of variables (36, 43–45).

In conclusion, this study revealed that three SNPs in the *AKT1* gene, alcohol intake ≥ 1 g/day, and BMI were the most influential variables for predicting breast cancer risk. While single genetic variants may not be enough to influence the risk, they may work together and interact with lifestyle factors (BMI and alcohol) to increase breast cancer risk. Our results provide insight into gene–lifestyle interactions and allow researchers to target efforts to promote intervention strategies to those within the population with risk genotypes. It also suggests the careful use of data on potential genetic targets in the intervention and clinical trials for cancer prevention to reduce the risk for breast cancer in postmenopausal women.

Jung et al.

Table 3. Combined effect of risk genotypes of *AKT1* rs2494740, *AKT1* rs2494744, and *AKT1* rs2498789 and dietary alcohol intake on breast cancer risk

	Total			BMI < 30 kg/m²			BMI ≥ 30 kg/m²		
n[a]	HR[b] (95% CI)	P	n	HR[b] (95% CI)	P	n	HR[b] (95% CI)	P
Risk genotypes									
0	reference		4,544	reference		1,969	**1.55 (1.22–1.96)**	<0.001
1+	1.89 (0.78–4.58)	0.161	35	0.66 (0.09–4.76)	0.685	19	**5.65 (2.08–15.36)**	<0.001
Dietary alcohol intake per day (g)									
0	reference		2,017	reference		1,087	**1.41 (1.01–1.97)**	0.044
1	1.15 (0.92–1.42)	0.213	2,562	1.05 (0.80–1.39)	0.709	901	**1.79 (1.30–2.47)**	<0.001
Risk genotypes combined with dietary alcohol intake per day									
0	reference		2,002	reference		1,078	1.39 (0.99–1.95)	0.056
1	1.13 (0.91–1.41)	0.267	2,557	1.04 (0.79–1.37)	0.788	900	**1.73 (1.25–2.40)**	<0.001
2	**2.83 (1.04–7.66)**	**0.040**	20	1.19 (0.16–8.55)	0.865	10	**7.10 (2.23–22.66)**	<0.001

Abbreviations: BMI, body mass index; CI, confidence interval; HR, hazard ratio. Numbers in bold face are statistically significant.
[a]The number of risk genotypes (*AKT1* rs2494740 AA, *AKT1* rs2494744 AA, and *AKT1* rs2498789 GG) defined as 0 (none) and 1+ (one or more of the alleles); the number of dietary alcohol intake defined as 0 (less than 1 g per day) and 1 (1 g or more per day); the number of combined risk genotypes and dietary alcohol intake defined as 0 (none), 1 (either risk genotypes or alcohol), and 2 (both).
[b]Multivariate regression was adjusted by age, family income, family history of breast cancer, depressive symptom, dietary alcohol per day (in risk genotypes analysis), percent calories from fat, percent calories from saturated fatty acids, physical activity, smoking, BMI (in total analysis), waist-to-hip ratio, age at menarche, age at menopause, pregnancy history, history of either hysterectomy or oophorectomy, and exogenous (unopposed and opposed) estrogen use.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: S.Y. Jung, J.C. Papp
Development of methodology: S.Y. Jung, J.C. Papp, E.M. Sobel
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.Y. Jung
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.Y. Jung, J.C. Papp, E.M. Sobel, Z.-F. Zhang
Writing, review, and/or revision of the manuscript: S.Y. Jung, J.C. Papp, E.M. Sobel, Z.-F. Zhang
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.Y. Jung, J.C. Papp
Study supervision: S.Y. Jung
Other (mentor of the first author): Z.-F. Zhang

Acknowledgments

Program Office: National Heart, Lung, and Blood Institute, Bethesda, MD: Jacques Rossouw, Shari Ludlam, Dale Burwen, Joan McGowan, Leslie Ford, and Nancy Geller.
Clinical Coordinating Center: Fred Hutchinson Cancer Research Center, Seattle, WA: Garnet Anderson, Ross Prentice, Andrea LaCroix, and Charles Kooperberg.
Investigators and Academic Centers: Brigham and Women's Hospital, Harvard Medical School, Boston, MA: JoAnn E. Manson; MedStar Health Research Institute/Howard University, Washington, DC: Barbara V. Howard; Stanford Prevention Research Center, Stanford, CA: Marcia L. Stefanick; The Ohio State University, Columbus, OH: Rebecca Jackson; University of Arizona, Tucson/Phoenix, AZ: Cynthia A. Thomson; University at Buffalo, Buffalo, NY: Jean Wactawski-Wende; University of Florida, Gainesville/Jacksonville, FL: Marian Limacher; University of Iowa, Iowa City/Davenport, IA: Robert Wallace; University of Pittsburgh, Pittsburgh, PA: Lewis Kuller; Wake Forest University School of Medicine, Winston-Salem, NC: Sally Shumaker.
Women's Health Initiative Memory Study: Wake Forest University School of Medicine, Winston-Salem, NC: Sally Shumaker.

References

1. American Cancer Society. Cancer fact and figures, 2017. Atlanta: American Cancer Society, Inc.: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2017/cancer-facts-and-figures-2017.pdf.
2. American Cancer Society. Colorectal cancer facts & figures 2014–2016. American Cancer Society, Inc. 2014: http://www.cancer.org/acs/groups/content/documents/document/acspc-042280.pdf.
3. Key TJ. Endogenous oestrogens and breast cancer risk in premenopausal and postmenopausal women. Steroids 2011;76:812–5.
4. Rose DP, Vona-Davis L. The cellular and molecular mechanisms by which insulin influences breast cancer risk and progression. Endocrine-related Cancer 2012;19:R225–41.
5. Weichhaus M, Broom J, Wahle K, Bermano G. A novel role for insulin resistance in the connection between obesity and postmenopausal breast cancer. Int J Oncol 2012;41:745–52.
6. Muendlein A, Lang AH, Geller-Rhomberg S, Winder T, Gasser K, Drexel H, et al. Association of a common genetic variant of the IGF-1 gene with event-free survival in patients with HER2-positive breast cancer. J Cancer Res Clin Oncol 2013;139:491–8.
7. Agurs-Collins T, Adams-Campbell LL, Kim KS, Cullen KJ. Insulin-like growth factor-1 and breast cancer risk in postmenopausal African-American women. Cancer Detect Prev 2000;24:199–206.
8. Li BD, Khosravi MJ, Berkel HJ, Diamandi A, Dayton MA, Smith M, et al. Free insulin-like growth factor-I and breast cancer risk. Int J Cancer 2001;91:736–9.
9. Ferroni P, Riondino S, Laudisi A, Portarena I, Formica V, Alessandroni J, et al. Pretreatment insulin levels as a prognostic factor for breast cancer progression. Oncologist 2016;21:1041–9.
10. Sieri S, Muti P, Claudia A, Berrino F, Pala V, Grioni S, et al. Prospective study on the role of glucose metabolism in breast cancer occurrence. Int J Cancer 2012;130:921–9.

11. Kabat GC, Kim M, Caan BJ, Chlebowski RT, Gunter MJ, Ho GY, et al. Repeated measures of serum glucose and insulin in relation to postmenopausal breast cancer. Int J Cancer 2009;125:2704–10.

12. Gunter MJ, Hoover DR, Yu H, Wassertheil-Smoller S, Rohan TE, Manson JE, et al. Insulin, insulin-like growth factor-I, and risk of breast cancer in postmenopausal women. J Nat Cancer Inst 2009;101:48–60.

13. Parekh N, Guffanti G, Lin Y, Ochs-Balcom HM, Makarem N, Hayes R. Insulin receptor variants and obesity-related cancers in the Framingham Heart Study. Cancer Causes Control 2015;26:1189–95.

14. Argiles JM, Lopez-Soriano FJ. Insulin and cancer (Review). Int J Oncol 2001;18:683–7.

15. Arcidiacono B, Iiritano S, Nocera A, Possidente K, Nevolo MT, Ventura V, et al. Insulin resistance and cancer risk: an overview of the pathogenetic mechanisms. Exp Diabetes Res 2012;2012:789174.

16. Al-Ajmi K, Ganguly SS, Al-Ajmi A, Mandhari ZA, Al-Moundhri MS. Insulin-like growth factor 1 gene polymorphism and breast cancer risk among arab omani women: a case-control study. Breast Cancer 2012; 6:103–12.

17. Al-Zahrani A, Sandhu MS, Luben RN, Thompson D, Baynes C, Pooley KA, et al. IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer. Hum Mol Genet 2006;15:1–10.

18. Slattery ML, Sweeney C, Wolff R, Herrick J, Baumgartner K, Giuliano A, et al. Genetic variation in IGF1, IGFBP3, IRS1, IRS2 and risk of breast cancer in women living in Southwestern United States. Breast Cancer Res Treat 2007;104:197–209.

19. Wang Q, Liu L, Li H, McCullough LE, Qi YN, Li JY, et al. Genetic and dietary determinants of insulin-like growth factor (IGF)-1 and IGF binding protein (BP)-3 levels among Chinese women. PloS One 2014;9:e108934.

20. Quan H, Tang H, Fang L, Bi J, Liu Y, Li H. IGF1(CA)19 and IGFBP-3-202A/ C gene polymorphism and cancer risk: a meta-analysis. Cell Biochem Biophys 2014;69:169–78.

21. Haiman CA, Han Y, Feng Y, Xia L, Hsu C, Sheng X, et al. Genome-wide testing of putative functional exonic variants in relationship with breast and prostate cancer risk in a multiethnic population. PLoS Genet 2013;9: e1003419.

22. Zhang H, Wang A, Ma H, Xu Y. Association between insulin receptor substrate 1 Gly972Arg polymorphism and cancer risk. Tumour Biol 2013;34:2929–36.

23. Slattery ML, Lundgreen A, John EM, Torres-Mejia G, Hines L, Giuliano AR, et al. MAPK genes interact with diet and lifestyle factors to alter risk of breast cancer: the Breast Cancer Health Disparities Study. Nutr Cancer 2015;67:292–304.

24. Henley SJ, Kanny D, Roland KB, Grossman M, Peaker B, Liu Y, et al. Alcohol control efforts in comprehensive cancer control plans and alcohol use among adults in the USA. Alcohol Alcohol 2014;49:661–7.

25. Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Bouvard V, et al. Carcinogenicity of alcoholic beverages. Lancet Oncol 2007;8:292–3.

26. Mogensen UB, Ishwaran H, Gerds TA. Evaluating random forests for survival analysis using prediction error curves. J Stat Soft 2012;50:1–23.

27. Hamidi O, Poorolajal J, Farhadian M, Tapak L. Identifying important risk factors for survival in kidney graft failure patients using random survival forests. Iranian J Public Health 2016;45:27–33.

28. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study.

The Women's Health Initiative Study Group. Controlled Clin Trials 1998;19:61–109.

29. Pechlivanis S, Wagner K, Chang-Claude J, Hoffmeister M, Brenner H, Forsti A. Polymorphisms in the insulin like growth factor 1 and IGF binding protein 3 genes and risk of colorectal cancer. Cancer Detect Prev 2007; 31:408–16.

30. National Cancer Institute. SEER Program: Comparative Staging Guide For Cancer. June 1993.

31. Cleveland RJ, Gammon MD, Edmiston SN, Teitelbaum SL, Britton JA, Terry MB, et al. IGF1 CA repeat polymorphisms, lifestyle factors and breast cancer risk in the Long Island Breast Cancer Study Project. Carcinogenesis 2006;27:758–65.

32. Wang Y, McCullough ML, Stevens VL, Rodriguez C, Jacobs EJ, Teras LR, et al. Nested case-control study of energy regulation candidate gene single nucleotide polymorphisms and breast cancer. Anticancer Res 2007;27: 589–93.

33. Yarden RI, Friedman E, Metsuyanim S, Olender T, Ben-Asher E, Papa MZ. Single-nucleotide polymorphisms in the p53 pathway genes modify cancer risk in BRCA1 and BRCA2 carriers of Jewish–Ashkenazi descent. Mol Carcinog 2010;49:545–55.

34. 1000 Genomes Browser Orientation. 2011.

35. Ishwaran H, Kogalur UB. Random survival forests for R. 2007. https://pdfs. semanticscholar.org/951a/84f0176076fb6786fdf43320e8b27094dcfa. pdf.

36. Chung RH, Chen YE. A two-stage random forest-based pathway analysis method. PloS One 2012;7:e36662.

37. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests 2008;2:841–860. https://projecteuclid.org/download/pdfview_1/ euclid.aoas/1223908043.

38. Inuzuka R, Diller GP, Borgia F, Benson L, Tay EL, Alonso-Gonzalez R, et al. Comprehensive use of cardiopulmonary exercise testing identifies adults with congenital heart disease at increased mortality risk in the medium term. Circulation 2012;125:250–59.

39. Bergman D, Halje M, Nordin M, Engstrom W. Insulin-like growth factor 2 in development and disease: a mini-review. Gerontology 2013;59:240–49.

40. Mitsiades CS, Mitsiades N, Koutsilieris M. The Akt pathway: molecular targets for anti-cancer drug development. Curr Cancer Drug Targets 2004;4:235–56.

41. Nicholson KM, Anderson NG. The protein kinase B/Akt signalling pathway in human malignancy. Cell Signal 2002;14:381–95.

42. Pelucchi C, Tramacere I, Boffetta P, Negri E, La Vecchia C. Alcohol consumption and cancer risk. Nutr Cancer 2011;63:983–90.

43. Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. Technol Health Care 2016;24:31–42.

44. Pang H, Lin A, Holford M, Enerson BE, Lu B, Lawton MP, et al. Pathway analysis using random forests classification and regression. Bioinformatics 2006;22:2028–36.

45. Chang JS, Yeh RF, Wiencke JK, Wiemels JL, Smirnov I, Pico AR, et al. Pathway analysis of single-nucleotide polymorphisms potentially associated with glioblastoma multiforme susceptibility using random forests. Cancer Epidemiol Biomarkers Prev 2008;17:1368–73.

46. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007;39:1423–34.

Cancer Prevention Research

Genetic Variants in Metabolic Signaling Pathways and Their Interaction with Lifestyle Factors on Breast Cancer Risk: A Random Survival Forest Analysis

Su Yon Jung, Jeanette C. Papp, Eric M. Sobel, et al.

Cancer Prev Res 2018;11:44-51. Published OnlineFirst October 26, 2017.

Updated version	Access the most recent version of this article at: doi:10.1158/1940-6207.CAPR-17-0143
Supplementary Material	Access the most recent supplemental material at: http://cancerpreventionresearch.aacrjournals.org/content/suppl/2017/12/08/1940-6207.CAPR-17-0143.DC1

Cited articles	This article cites 40 articles, 5 of which you can access for free at: http://cancerpreventionresearch.aacrjournals.org/content/11/1/44.full#ref-list-1

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cancerpreventionresearch.aacrjournals.org/content/11/1/44. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.