Title

Genetic variants of glucose metabolism and exposure to smoking in African American breast cancer.

Journal

Endocrine Related Cancer, 30(4)

ISSN

1351-0088

Authors

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

Publication Date

2023-04-01

DOI

10.1530/erc-22-0184

Peer reviewed

HHS Public Access

Author manuscript

Endocr Relat Cancer. Author manuscript; available in PMC 2023 April 12.

Published in final edited form as:

Endocr Relat Cancer. 2023 April 01; 30(4): . doi:10.1530/ERC-22-0184.

Genetic variants of glucose metabolism and exposure to smoking in African American breast cancer

Su Yon Jung[1,2,*], **Jeanette C. Papp**[2,3], **Eric M. Sobel**[3,4], **Matteo Pellegrini**[5], **Herbert Yu**[6]

[1]Translational Sciences Section, School of Nursing, University of California, Los Angeles, Los Angeles, CA 90095, USA

[2]Jonsson Comprehensive Cancer Center, University of California, Los Angeles, Los Angeles, CA 90095, USA

[3]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA

[4]Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA

[5]Department of Molecular, Cell and Developmental Biology, Life Sciences Division, University of California, Los Angeles, Los Angeles, CA 90095, USA

[6]Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI 96813, USA

Abstract

Insulin resistance (IR) is a well-established risk factor for breast cancer (BC) development in African American (AA) postmenopausal women. While obesity and IR are more prevalent in AA than white women, they are under-represented in genome-wide studies for systemic regulation of IR. By examining 780 genome-wide IR single-nucleotide polymorphisms (SNPs) available in our data, we tested 4,689 AA women in a Random Survival Forest framework. With 37 BC-associated lifestyle factors, we conducted a gene–environment interaction analysis to estimate risk prediction for BC with the most influential genetic and behavioral factors and evaluated their combined

[*]**Correspondence:** Su Yon Jung, Ph.D., M.P.H., Associate Professor for Translational Sciences Section, Jonsson Comprehensive Cancer Center, School of Nursing, University of California Los Angeles, 700 Tiverton Ave, 3-264 Factor Building, Los Angeles, CA 90095, USA, Phone: (310) 825-2840, Fax: (310) 267-0413, sjung@sonnet.ucla.edu.

Authors' contributions

SYJ, JP, ES, MP, and HY designed the study. SYJ performed the genomic data QC and the statistical analysis and interpreted the data. JP and ES supervised the genomic data QC and analysis. MP and HY participated in the study coordination and interpreting the data. SYJ secured funding for this project. HY supervised the project. All participated in the paper writing and editing. All authors have read and approved the submission of the manuscript.

Conflict of interest: All authors declare no potential conflict of interest.

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: JoAnn E. Manson, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; Barbara V. Howard, MedStar Health Research Institute/Howard University, Washington, DC; Marcia L. Stefanick, Stanford Prevention Research Center, Stanford, CA; Rebecca Jackson, The Ohio State University, Columbus, OH; Cynthia A. Thomson, University of Arizona, Tucson/Phoenix, AZ; Jean Wactawski-Wende, University at Buffalo, Buffalo, NY; Marian Limacher, University of Florida, Gainesville/Jacksonville, FL; Robert Wallace, University of Iowa, Iowa City/Davenport, IA; Lewis Kuller, University of Pittsburgh, Pittsburgh, PA; and Sally Shumaker, Wake Forest University School of Medicine, Winston-Salem, NC.

and joint effects on BC risk. By accounting for variations of individual SNPs in BC in the prediction model, we detected 4 fasting glucose–associated SNPs in *PCSK1*, *SPC25*, *ADCY5*, and *MTNR1B* and 3 lifestyle factors (smoking, oral contraceptive use, and age at menopause) as the most predictive markers for BC risk. Our joint analysis of risk genotypes and lifestyle with smoking revealed a synergistic effect on increased risk of BC, particularly ER/PR+ BC, in a gene–lifestyle dose-dependent manner. The joint effect of smoking was more substantial in women with a prolonged exposure to cigarette smoking and female hormones. The top GWA-SNPs associated with metabolic biomarkers in combination with lifestyles synergistically increase the predictability of invasive ER/PR positive BC risk among AA women. Our findings highlight generically targeted preventive interventions for women who carry particular risk genotypes and lifestyles.

Keywords

glucose homeostasis; random survival forest; smoking; oral contraceptive; breast cancer; African American postmenopausal women

Introduction

Breast cancer (BC) is the most common cancer diagnosis and cause of death related to cancer in women in the United States and worldwide; more than 80% of new cases and deaths occur in postmenopausal women ages 50 years and older. Whereas African American (AA) women have a lower BC incidence rate than white women (e.g., 126.7 vs. 130.8 cases per 100,000 during 2012–2016), AA women's incidence rates have more rapidly increased than those in white women (0.9% vs. 0.4% per year), contributing to a convergence in BC incidence rates in 2016. BC is also the top leading cause of cancer incidence and mortality in AA women.

Obesity and insulin resistance (IR) are well-established risk factors for postmenopausal BC (White et al., 2012, Robinson et al., 2014, Munsell et al., 2014, Dash et al., 2015, Capasso et al., 2010). In particular, obesity and lack of estrogen in postmenopausal women promote IR, leading to glucose intolerance, which is characterized by hyperglycemia and compensatory hyperinsulinemia. IR has a mitotic effect by overexpression of insulin receptors (INRs) and insulin-like growth factor 1 receptors (IGF1Rs) and dysregulation of downstream cellular insulin-signaling cascades, resulting in enhancement of cellular anabolic status and anti-apoptosis. IR may therefore initiate and progress cancer cell growth in breast tissues of postmenopausal women (Strange et al., 2004, Becker et al., 2009). Obesity and IR are more prevalent in AA women (Gallagher et al., 2016, Samson et al., 2016), indicating that they are more likely to be metabolically unhealthy than white women are. Accordingly, BC cells and tissues of AA women, compared with those of white women, show higher expressions of *Resistin* (a gene connected to obesity and IR) (Stewart et al., 2013, Vallega et al., 2016) and INR (Gallagher et al., 2016, Kalla Singh et al., 2011); they also present higher IGF1R and insulin receptor substrate 1 phosphorylation (Kalla Singh et al., 2010) and miRNA involvement in the interconnected insulin signaling pathways (Sugita et al., 2016), implicating the critical role of IR in AA women's BC development.

For investigating the systemic regulation of IR, extensive relevant genomic studies have been conducted, but mostly focusing on whites; racial/ethnic minorities, particularly AAs, are under-represented in the genomic study of IR phenotypes at the genome-wide level. Discovering IR-genetic signatures associated with AAs can advance the understanding of mechanisms related to IR at the molecular level in the AA population and as potential risk factors of BC, improve accuracy in predicting BC development, thus contributing to the development of genetically targeted personalized preventive/therapeutic interventions. In addition, previous genomic studies of IR explained a small to moderate proportion of BC variation (Jung et al., 2019), requiring the involvement of lifestyle factors in gene–environment (G×E) interaction studies to generate risk profiles for BC that combine both genes and lifestyles.

For those reasons, we conducted a genomic G×E study focusing on AA postmenopausal women. By examining 780 genome-wide association (GWA)–based IR single-nucleotide polymorphisms (SNPs), we first performed a validation test with the relevant phenotype in our dataset. Next, we identified BC-associated lifestyle factors from the literature that were disproportionately represented in AA women, and incorporated those lifestyle factors with the validated genetic markers to estimate risk prediction for BC in a Random Survival Forest (RSF) prediction model. We eventually constructed BC risk profiles with the most predictive genetic and behavioral factors and evaluated their combined and joint effects on BC risk. We hoped that this multimodal analytic approach could provide a potential way to resolve the inconclusive findings from previous studies of genetic and lifestyle factors in association with BC and improve the predictive ability for BC risk in the racial/ethnic minority AA women.

Materials and Methods

Study population

Our study included AA postmenopausal women enrolled in the Women's Health Initiative Database for Genotypes and Phenotypes (WHI dbGaP) SNP Health Association Resource (SHARe) for the minorities AAs and Hispanics, which is part of the WHI Harmonized and Imputed GWA Studies (GWASs) with a joint imputation and harmonization effort for many of the GWASs within the WHI Study. Details of study designs and rationale have been reported (2019, 1998, 2021d). Healthy women were enrolled in the WHI study between 1993 and 1998 at 40 clinical centers across the U.S. if they were 50–79 years old, postmenopausal, and able to provide written informed consent. Participants were further screened for the WHI dbGaP study if they had met eligibility requirements for data submission to dbGaP and provided DNA samples. Among a total of 7,470 women reporting their race or ethnicity as AA, we applied exclusion criteria (genomic data quality control [QC]; history of diabetes; < 1-year follow-up; and diagnosis of any cancer type at screening), resulting in 4,689 AA women. They had been followed up through August, 2014, with a 15-year median follow-up; by their last follow-up, 271 (5%) of whom had developed primary invasive BC. This study was approved by the institutional review boards of each WHI participating clinical center and the University of California, Los Angeles.

Selection of IR genetic variants

We used data available on glycemic traits from the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC, www.magicinvestigators.org) (Dupuis et al., 2010, Scott et al., 2012, Manning et al., 2012, Lagou et al., 2021), in which fasting levels of glucose (FG) and insulin (FI) were analyzed as continuous variables. We also used 2 other GWAS resources examining an AA cohort: one (Ramos et al., 2011) determined query SNPs in a 500-kb linkage disequilibrium (LD) block in association with FG, and the other (Mondal et al., 2013) investigated functional SNPs for glucose homeostasis. Among a total of 1,344 FG-SNPs and 313 FI-SNPs, 689 FG-SNPs and 91 FI-SNPs are available in our AA SHARe genomic dataset, of which 94 FG-SNPs (34 index in LD < 0.3) and 8 FI-SNPs (4 index in LD < 0.3) were validated with a relevant phenotype.

Genotyping and phenotyping

Genotyping data specifically from AA women were extracted for this study from the WHI dbGaP SHARe. Details of the genotyping are found elsewhere (2019, 2021d). Briefly, DNA samples were obtained from participants' blood samples at baseline and genotyped via Affymetrix 6.0 (Affymetrix, Inc., Santa Clara, CA). Data normalization to Genome Reference Consortium Human Build 37, imputation with the 1000 genomes refence panels, and harmonization with pairwise concordance among all samples across WHI GWASs were performed at the Fred Hutchinson Cancer Research Center in Seattle, WA. We conducted data QC, filtering on SNPs with a missing-call rate of < 2%, a Hardy-Weinberg equilibrium of $p \geq 1E{-}04$, and $\widehat{R}^2 \geq 0.6$ imputation quality (Schumacher et al., 2018). Further, we excluded individuals with unexpected duplicates, first- and second-degree relatives, and outliers on the basis of genetic principal components (PCs).

Fasting blood samples had been obtained from each participant at baseline by trained phlebotomists. Serum levels of glucose and insulin had been measured via the hexokinase method on a Hitachi 747 instrument (Boehringer Mannheim Diagnostics, Indianapolis, IN) and by a radioimmunoassay method (Linco Research, Inc., St. Louis, MO), respectively, with average coefficients of variation of 1.28% and 10.93%, respectively.

Lifestyle variables and BC outcome

We initially conducted a literature review (Akinyemiju et al., 2018, Williams et al., 2016, Williams et al., 2017, Boggs et al., 2015, Colonna et al., 2012, Xing et al., 2020, Park et al., 2016, Butler et al., 2016) on lifestyle factors that are associated with BC in AA women and extracted lifestyle variables: nutrition (daily fruits and vegetables and dietary fiber); physical activity; alcohol intake (dietary alcohol/day and history of alcohol intake); smoking (years as a regular smoker and number of cigarettes/day); family history of breast cancer (genetic inheritance); body mass index (BMI) and abdominal adiposity (waist circumference and waist-to-hip ratio); comorbid conditions (lipid metabolic profile [hypercholesterolemia ever], cardiovascular disease ever, hypertension ever, and depressive symptoms); duration of oral contraceptive (OC) use; ages at menarche and menopause, number of pregnancies; breast feeding period; oophorectomy and hysterectomy; and exogenous estrogen (E) use (E only and E plus progestin [P]). In addition to those variables, we incorporated in our

data analysis demographic and socioeconomic factors, such as age at enrollment, education, employment, marital status, and other dietary variables, including dietary calcium, vitamin K, total sugars, and percent calories from saturated fatty acid (SFA), mono- and poly-unsaturated FA, carbohydrates, and protein. All the participants' variables were obtained at screening by self-administered questionnaires (except height, weight, and waist/hip circumferences, which were measured by trained staff), and the data collection process was monitored by the coordinating clinical centers. With a total of 37 variables, we performed preliminary univariate and stepwise multiple regression analyses for BC risk and checked multicollinearity among variables.

Primary invasive BC development and its time to develop were assessed. In detail, Participants' BC outcomes were determined via a centralized review of medical records and pathology and cytology reports by the WHI committee of physicians and were coded according to the National Cancer Institute's Surveillance, Epidemiology, and End-Results guidelines (National Cancer Institute, June 1993). The time between study enrollment and BC diagnosis, censoring, or study end-point was estimated as the number of days and then converted into years.

Statistical analysis

The associations of GWA IR-SNPs with naturally log-transformed FG (mg/dl) and FI (μIU/ml) and with BC risk were estimated via linear and Cox proportional hazards regressions, respectively, both of which were adjusted for age and 10 genetic PCs with assumptions for each regression met. A 2-tailed p value < 0.05 for validation tests with FG and FI was considered nominally significant, and after the Bonferroni correction for multiple comparisons, p < 7E-05 for FG and < 5E-04 for FI were considered statistically significant.

With index as well as individual SNPs and lifestyle variables, we performed RSF, a nonparametric tree-based ensemble learning method that accounts for the nonlinear effects and high-order interactions among variables (Mogensen et al., 2012). RSF has outperformed traditional prediction models and successfully yielded accurate predictions (Chung and Chen, 2012, Montazeri and Beigzadeh, 2016, Pang et al., 2006, Chang et al., 2008, Tong et al., 2018). Detailed descriptions for RSF are found in Table S1. Briefly, the process to generate a tree from each bootstrapped sample to maximize risk differences across daughter nodes was repeated numerous times (n = 50,000 trees in this study) to create a forest of trees (Ishwaran and Kogalur, 2007, Chung and Chen, 2012). The prediction error (i.e., misclassification probability) was estimated using the out-of-bag (OOB) data to calculate the OOB concordance index (c-index = 1 – prediction error). The OOB c-index is a quantitative measure of prediction performance, conceptually similar to the area under the receiver operating characteristic (AUROC) curve (Ishwaran and Kogalur, 2007, Ishwaran et al., 2008). The predictability of each variable was determined with 2 values: a) minimal depth (MD), in which variables with a small MD are considered highly predictive, and b) variable importance (VIMP), calculated from the permutation of OOB datasets, in which variables with a larger VIMP are more predictive (Inuzuka et al., 2012, Mogensen et al., 2012).

We conducted a 2-stage RSF analysis: SNPs and lifestyle factors separately in the first RSF (Figure S1) and then with only SNPs and lifestyle factors with significantly low MD and high VIMP values carried over to the second RSF. This approach enabled us to exclude variables without significant effects on BC, thus giving more statistical power with the correct type I error in the second stage. In the second RSF, we took a multimodal approach to detect the most predictive variables: a) comparing the MD and VIMP measures in the plot; b) estimating OOB c-index with the nested RSF model; and c) estimating the incremental error rate of each variable in the nested sequence of RSF models. Finally, we estimated the individual and combined effects of the most influential risk genotypes and lifestyle factors on BC risk via multiple Cox regression adjusted for covariates. A 2-tailed p value was corrected for multiple comparisons by the Benjamini-Hochberg method; a 5% false discovery rate (FDR) was considered statistically significant. R4.0.4 (survival, metaphor, forestplot, survivalROC, randomForestSRC, ggRandomForests, ggplot2, ggthemes, and gamlss) was used.

Results

Two-stage multimodal RSF

Among 94 FG-SNPs (Table S2A) and 8 FI-SNPs (Table S2B) validated in our data at significance nominally and after multiple comparison corrections, there are 34 index FG-SNPs and 4 index FI-SNPs. Considering that a combined analysis of index SNPs only can mask individual SNPs' variation in the risk of BC development, we examined all phenotype-specific individual SNPs, including index SNPs, in the RSF prediction model. In the first RSF (Figure S1), we generated a plot for lifestyle factors and SNPs separately to compare the 2 prediction measures MD and VIMP and identified the most influential factors in agreement with high ranks: 9 of 37 lifestyles, 5 of 34 FG index SNPs, 4 (1 of which overlaps with 1 of the selected FG indexes) of 94 FG individual SNPs, and 3 (including 2 indexes) of 8 FI SNPs.

In the second RSF, with the selected 9 lifestyles and 11 FG/FI SNPs, we first used the 2 measures MD and VIMP (Table 1) to generate a plot (Figure 1A). Both measures with high ranks detected 4 FG SNPs (*ADCY5* rs6798189; *MTNR1B* rs7945617; *SPC25* rs17539351; and *PCSK1* rs13169290) and 3 lifestyles (number of years as a regular smoker, duration of OC use, and age at menopause) as the most predictive variables for BC development. We also computed the c-index (= AUROC) from the nested RSF models and plotted it with variables ranked by MD (Figure 1B), detecting the same set of top 7 variables that are distinct to improve prediction accuracy compared with other variables; this suggests the complementary use of the c-index for determining variables' prediction ability. Further, we computed a dropping error in each variable in the nested sequence of RSF models (Table 1), determining once again the same top 7 variables as the strongest contributors to reduce the prediction error rate.

Combined and joint effects of the most dominant IR SNPs and lifestyles on BC risk

Using the RSF method, which accounts for confounding factors and nonlinearity effect of each variable on BC outcomes, we computed the predicted cumulative BC incidence

rate for the top 7 variables (Figure 2). Accordingly, the risk genotype of each SNP was categorized for further analysis as follows (Figures 2A–2D): *ADCY5* rs6798189 AG+GG; *MTNR1B* rs7945617 CC; *SPC25* rs17539351 TC+CC; and *PCSK1* rs13169290 AG+GG. Corresponding to the cutoff values in Figures 2E–2G, risk lifestyles were defined as ≥ 20 years of regular smoking; ≤ 5 years of OC use; and age > 45 years at menopause. For those genotypes and lifestyles, individually and combined, the risk for BC was estimated, with HRs ranging from 1.19 to 2.13 for SNPs (Table S3) and from 1.19 to 2.21 for lifestyles (Table S4), confirming the single and combined effects of genetic and lifestyle factors on BC risk.

However, when those genetic and lifestyle factors, singly and in combination, were tested for the joint effect with smoking that incorporates age at menopause, much stronger risks for BC were observed (Tables 2–5). In detail, throughout the analyses, we conducted a stratification analysis by age at menopause (≤ 45 vs. > 45 years), where an earlier menopause group (EMG) is more likely than a later menopause group (LMG) to undergo artificial menopause as indicative of > 80% with a history of hysterectomy and/or oophorectomy. First, SNPs and lifestyles were combined, each without consideration of smoking, and compared their risk for BC between menopause groups (the first columns in Tables 2 and 4), revealing a greater risk for BC of SNPs or lifestyles in the LMG than in the EMG. Similarly, the SNPs and lifestyles, when combined, presented a more profound increased risk of BC in the LMG (Table 4; $HR_{genotype\ plus\ behaviors}$ = 4.01, 95% CI: 1.85–8.71) than in the EMG (Table 2; $HR_{genotype\ plus\ behaviors}$ = 1.95, 95% CI: 0.91–4.20). This indicates a prolonged female hormonal effect on FG SNPs and lifestyles led to increased BC risk.

Next, we examined the risk genotypes and lifestyle (e.g., OC use) to determine their BC risk jointly with smoking in each menopause group. Smokers in each group were categorized to compare between never smokers and regular smokers ≥ 20 years and between shorter- and longer-term regular smokers (< 20 vs. ≥ 20 years). Overall, in both menopause groups, the risk difference (i.e., the joint effect of smoking) in BC between never smokers and regular smokers for ≥ 20 years who carry risk genotypes and lifestyle, separately and together, was substantially greater than the difference in BC risk between regular smokers (< 20 vs. ≥ 20 years) who carry genetic and lifestyle factors, singly and in combination. For example, in the LMG, compared with never smokers who have null-risk genotypes and lifestyle (e.g., > 5 years of OC use), regular smokers for ≥ 20 years who have both risk genotypes and lifestyle (e.g., ≤ 5 years of OC use) had almost 9 times increased risk for BC (Table 4). In contrast, no significant difference in BC risk was detected between shorter- and longer-term regular smokers (< 20 years vs. ≥ 20 years) who carry the SNPs and lifestyle (Table 5). Similar patterns were observed in the EMG (Tables 2 and 3). This indicates a strong smoking effect on FG SNPs and lifestyle to increase BC risk across different lengths of accumulated exposure to female hormones.

Further, we compared the joint effect of smoking on BC risk between the EMG and the LMG by focusing on risk genotypes and OC use separately. The joint effects of never smokers vs. regular smokers for ≥ 20 years who carry the SNPs or used OC for ≤ 5 years did not differ between menopause groups (Tables 2 and 4). On the contrary, the joint effects

of regular smoking for < 20 years vs. ≥ 20 years on AA women who carry SNPs or used OC for ≤ 5 years were higher in the LMG (Table 5) than in the EMG (Table 3). This implies that a substantial difference in smoking exposure (i.e., never vs. longer-term smokers) may mask the effect of female hormones on BC risk, but greater accumulated hormones can exert their effects more apparently among regular smokers (< 20 vs. ≥ 20 years) when a relatively smaller difference in smoking exposure period exists.

We further conducted BC molecular subtype–specific analyses for the combined and joint effects of SNPs and lifestyles with smoking by menopause status (Table S5), yielding similar patterns but a substantially increased effect on the risk for estrogen/progesterone positive (ER/PR+) BC.

Discussion

Our genetic G×E study in postmenopausal AA women examined an extensive set of GWA IR-SNPs to test for interactions with lifestyle factors in the RSF BC risk–prediction model. By accounting for variations of individual SNPs in BC risk, we eventually detected 4 FG SNPs (including 2 indexes) and 3 lifestyle factors as the most predictive variables for BC risk. Constructing BC risk profiles that incorporate genetic markers, phenotypes, and behaviors seemed to improve cancer prediction, leading to the promotion of gene–lifestyle combined interventions for cancer prevention effort. Our joint analysis of the risk genotypes and smoking revealed a synergistic effect on increased risk of BC, particularly for ER/PR+ BC, in a gene–lifestyle dose-dependent manner, indicating a substantial improvement in risk prediction for BC.

All 4 of the FG SNPs are located in the introns of genes that play well-established roles in regulating insulin secretion and sensitivity, postulating that their genetic variation may affect glucose metabolism. For example, the *PCSK1* gene is 1 of the genes linked to early-onset obesity. Encoded prohormone convertase 1/3 is involved in the biosynthetic process of prohormones in endocrine tissues, regulating food ingestion and glucose homeostasis (Ramos-Molina et al., 2016, 2021e). Also, the *SPC25* gene involved in kinetochore–microtubule interaction was reported to regulate FG and BMI (2021b) and, in particular, play a key role in DNA damage repair and kinetochore assembly, thus promoting proliferation of BC cells (Wang et al., 2019, Pathania et al., 2016). Both genes' variants in our study exhibited effects on glucose homeostasis and an increased risk for BC despite insufficient power.

The *ADCY5* gene encodes adenylate cyclase 5, catalyzing the production of cyclic adenosine monophosphate that is the second messenger involved in the stimulation of pancreatic cells' insulin secretion (Roman et al., 2017, Lin et al., 2020, 2021c). Genetic variants in this gene may result in impaired insulin secretion, thus the aberrant response to glucose metabolism. One of the *ADCY5* gene networks is involved in BC regulation by interacting with *Stathmin1*, which plays an important role in carcinogenesis (2021a, Obayashi et al., 2017). Our study is the first to detect the association of this gene's variation with BC risk. Also, *MTNR1B* is a well-known gene that encodes melatonin receptor 1B, which plays a key role in insulin secretion and glucose metabolism in pancreatic islets

(McMullan et al., 2013, Stumpf et al., 2008). Its genetic variants may disturb circadian rhythm, resulting in glucose intolerance (Hu and Jia, 2016). In vitro studies of elevated expression of this gene found mixed results for BC and other cancers: a protective effect by inhibiting BC ERα mRNA expression (Yuan et al., 2002) or a cancer-progression effect in other cancers (Liu et al., 2019, Leon et al., 2012). Genetic variation of this gene in our study, consistent with the results in a previous genetic study (Deming et al., 2012), was associated with BC risk. Those particular genotypes of SNPs that are associated in our study with BC risk in *ACDY5* and *MTNR1B* are required for further functional studies.

Given that premenopausal hysterectomy and oophorectomy are associated with a lower risk of BC, while age at menopause of 45 years or older is associated with a higher risk of BC than earlier menopause (Cui et al., 2014), we examined the combined and joint effects of SNPs and lifestyles among the earlier (80% of women with removal of ovaries and/or uterus) and later menopause groups, separately. As expected, the FG SNPs and lifestyles displayed a greater risk in the LMG than in the EMG. When smoking was incorporated, this longer-term accumulated effect of female hormone exposure on BC risk was constant, but only among regular smokers (< 20 vs. ≥ 20 years); no difference in risk of BC between the menopause groups was observed in the comparison of never and regular smokers. This suggests that long-term smoking effects on BC carcinogenesis may override female hormonal modification on BC risk.

Higher total nicotine equivalents after controlling for number of cigarettes per day have been reported in AAs than in other races/ethnicities (Park et al., 2015), but little data exist for AA women in association with BC. In particular, AA postmenopausal women have an increased risk of BC with a longer duration of smoking (≥ 20 years) (Park et al., 2016, Butler et al., 2016); this can be partially explained by the increased level of female hormones induced by cigarette smoking as well as by tobacco's direct toxic and carcinogenic effects in breast tissues (Key et al., 2011). Also, the effect of cigarette smoking on BC risk can be BC-subtype specific. In agreement with previous reports (Park et al., 2016, Butler et al., 2016), our results showed that the joint effect of smoking on BC risk is more substantial in ER/PR+ BC. In previous studies, the magnitude of the association between smoking and ER/PR+ BC was modest (HRs, 1.16–1.51). Notably, our study revealed that in both EMG and LMG, long-term smokers (≥ 20 years) with combined FG SNPs had an 8-times higher risk of ER/PR+ BC than never smokers without the SNPs, although our induction period, 15 years on average, was somewhat shorter than the typical induction period (Terry et al., 2002). This suggests that smoking and glucose metabolism are interrelatedly involved in the ER/PR+ BC carcinogenesis, warranting future biologic mechanism studies on the cigarette–IR interaction.

In our study, OC users for < 5 years had a greater risk of BC than OC users for ≥ 5 years. This finding is consistent with those of other studies (Trentham-Dietz et al., 2000, Urban et al., 2012) particularly among women who ceased OC use within 10 years. Although OCs are considered carcinogenic and increase the risk for BC (IARC, 2007), exogenous E plus P use increases the risk to a much greater extent than E-only use (Chlebowski et al., 2016, Rosenberg et al., 2016). Synthetic progestin has an affinity for androgen and mineralocorticoid receptors, leading to proliferation and anti-apoptosis in breast cells

(Cogliano et al., 2005a). Thus, the difference in risk for BC according to the length of use may be a result of changes in OC formulation. The amount of progestin in OC formulations has declined and new types of progestin have been introduced over time (Cogliano et al., 2005b), supporting our finding of reduced BC risk in longer-term OC users. Data were not available for our study on the recency and type of OC formulation, so future studies integrating those factors are warranted.

Our data on smoking and OC use were self-reported, so they could have been biased by misclassifications. However, several validation studies have shown the high reliability of those self-reported measures for active smoking assessment (Soulakova et al., 2012) and OC use (Nischan et al., 1993, Rosenberg et al., 1983). Also, a 2-stage RSF may overfit the model with multiple tasks, requesting a replication study for particularly, high order interactions, with an independent dataset. Our study focused on AA postmenopausal women, the generalizability of our findings to other populations is limited.

Overall, our study indicates that IR SNPs and OC use, jointly with smoking, synergistically increase the risk of BC, more substantially in ER/PR+ BC. The joint effect of smoking on ER/PR+ BC was more profound in women who had long-term exposure to cigarette smoking and a prolonged exposure to female hormones. Our findings may improve prediction accuracy in BC subtypes, and highlight genetically targeted preventive interventions (e.g., smoking cessation) for AA women who carry particular risk genotypes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Part of the data for this project was provided by the WHI program, which is funded by the National Heart, Lung, and Blood Institute, the National Institutes of Health, and the U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession (phs000200.v11.p3).

Funding

This study was supported by the National Institute of Nursing Research of the National Institutes of Health under Award Number K01NR017852.

References

American Cancer Society. Breast Cancer Facts & Figures 2019–2020 Atlanta: American Cancer Society, Inc. 2019.

American Cancer Society. Cancer Fact and Figures 2021 Atlanta: American Cancer Society, Inc. 2021.

American Cancer Society. Cancer Fact and Figures for African Americans 2019–2021 Atlanta: American Cancer Society, Inc. 2021.

1998. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control Clin Trials, 19, 61–109.

2019. NCBI: WHI Harmonized and Imputed GWAS Data. A sub-study of Women's Health Initiative [Online] Available: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000746.v3.p3.

2021a. Gene Card: Human Gene Database: ADCY5 Gene (Protein Coding) [Online] Available: https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADCY5 [Accessed].

2021b. Gene Card: Human Gene Database: SPC25 Gene (Protein Coding) [Online] Available: https://www.genecards.org/cgi-bin/carddisp.pl?gene=SPC25&keywords [Accessed].

2021c. Genetic Home Reference - ADCY5 gene [Online] Available: https://medlineplus.gov/genetics/gene/adcy5/ [Accessed].

2021d. NCBI: Women's Health Initiative - SNP Health Association Resource. A sub-study of Women's Health Initiative [Online] Available: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000386.v8.p3.

2021e. PCSK1 proprotein convertase subtilisin/kexin type 1 [Homo sapiens (human)] - Gene - NCBI [Online] Available: https://www.ncbi.nlm.nih.gov/gene/5122 [Accessed].

AKINYEMIJU T, MOORE JX & PISU M 2018. Mediating effects of cancer risk factors on the association between race and cancer incidence: analysis of the NIH-AARP Diet and Health Study. Ann Epidemiol, 28, 33–40 e2. [PubMed: 29217211]

BECKER S, DOSSUS L & KAAKS R 2009. Obesity related hyperinsulinaemia and hyperglycaemia and cancer development. Arch Physiol Biochem, 115, 86–96. [PubMed: 19485704]

BOGGS DA, ROSENBERG L, ADAMS-CAMPBELL LL & PALMER JR 2015. Prospective approach to breast cancer risk prediction in African American women: the black women's health study model. J Clin Oncol, 33, 1038–44. [PubMed: 25624428]

BUTLER EN, TSE CK, BELL ME, CONWAY K, OLSHAN AF & TROESTER MA 2016. Active smoking and risk of Luminal and Basal-like breast cancer subtypes in the Carolina Breast Cancer Study. Cancer Causes Control, 27, 775–86. [PubMed: 27153846]

CAPASSO I, ESPOSITO E, PENTIMALLI F, CRISPO A, MONTELLA M, GRIMALDI M, DE MARCO M, CAVALCANTI E, D'AIUTO M, FUCITO A, FRASCI G, MAUREA N, ESPOSITO G, PEDICINI T, VECCHIONE A, D'AIUTO G & GIORDANO A 2010. Metabolic syndrome affects breast cancer risk in postmenopausal women: National Cancer Institute of Naples experience. Cancer Biol Ther, 10, 1240–3. [PubMed: 20935521]

CHANG JS, YEH RF, WIENCKE JK, WIEMELS JL, SMIRNOV I, PICO AR, TIHAN T, PATOKA J, MIIKE R, SISON JD, RICE T & WRENSCH MR 2008. Pathway analysis of single-nucleotide polymorphisms potentially associated with glioblastoma multiforme susceptibility using random forests. Cancer Epidemiol Biomarkers Prev, 17, 1368–73. [PubMed: 18559551]

CHLEBOWSKI RT, ANDERSON GL, ARAGAKI AK & PRENTICE R 2016. Breast Cancer and Menopausal Hormone Therapy by Race/Ethnicity and Body Mass Index. J Natl Cancer Inst, 108.

CHUNG RH & CHEN YE 2012. A two-stage random forest-based pathway analysis method. PLoS One, 7, e36662. [PubMed: 22586488]

COGLIANO V, GROSSE Y, BAAN R, STRAIF K, SECRETAN B & EL GHISSASSI F 2005a. Carcinogenicity of combined oestrogen-progestagen contraceptives and menopausal treatment. Lancet Oncol, 6, 552–3. [PubMed: 16094770]

COGLIANO V, GROSSE Y, BAAN R, STRAIF K, SECRETAN B, EL GHISSASSI F & CANCER, W. H. O. I. A. F. R. O. 2005b. Carcinogenicity of combined oestrogen-progestagen contraceptives and menopausal treatment. Lancet Oncol, 6, 552–3. [PubMed: 16094770]

COLONNA SV, DOUGLAS CASE L & LAWRENCE JA 2012. A retrospective review of the metabolic syndrome in women diagnosed with breast cancer and correlation with estrogen receptor. Breast Cancer Res Treat, 131, 325–31. [PubMed: 21964613]

CUI Y, DEMING-HALVERSON SL, SHRUBSOLE MJ, BEEGHLY-FADIEL A, FAIR AM, SANDERSON M, SHU XO, KELLEY MC & ZHENG W 2014. Associations of hormone-related factors with breast cancer risk according to hormone receptor status among white and African American women. Clin Breast Cancer, 14, 417–25. [PubMed: 24970715]

DASH C, MAKAMBI K, WALLINGTON SF, SHEPPARD V, TAYLOR TR, HICKS JS & ADAMS-CAMPBELL LL 2015. An exercise trial targeting African-American women with metabolic syndrome and at high risk for breast cancer: Rationale, design, and methods. Contemp Clin Trials, 43, 33–8. [PubMed: 25962889]

DEMING SL, LU W, BEEGHLY-FADIEL A, ZHENG Y, CAI Q, LONG J, SHU XO, GAO YT & ZHENG W 2012. Melatonin pathway genes and breast cancer risk among Chinese women. Breast Cancer Res Treat, 132, 693–9. [PubMed: 22138747]

DUPUIS J, LANGENBERG C, PROKOPENKO I, SAXENA R, SORANZO N, JACKSON AU, WHEELER E, GLAZER NL, BOUATIA-NAJI N, GLOYN AL, LINDGREN CM, MAGI R, MORRIS AP, RANDALL J, JOHNSON T, ELLIOTT P, RYBIN D, THORLEIFSSON G, STEINTHORSDOTTIR V, HENNEMAN P, GRALLERT H, DEHGHAN A, HOTTENGA JJ, FRANKLIN CS, NAVARRO P, SONG K, GOEL A, PERRY JR, EGAN JM, LAJUNEN T, GRARUP N, SPARSO T, DONEY A, VOIGHT BF, STRINGHAM HM, LI M, KANONI S, SHRADER P, CAVALCANTI-PROENCA C, KUMARI M, QI L, TIMPSON NJ, GIEGER C, ZABENA C, ROCHELEAU G, INGELSSON E, AN P, O'CONNELL J, LUAN J, ELLIOTT A, MCCARROLL SA, PAYNE F, ROCCASECCA RM, PATTOU F, SETHUPATHY P, ARDLIE K, ARIYUREK Y, BALKAU B, BARTER P, BEILBY JP, BEN-SHLOMO Y, BENEDIKTSSON R, BENNETT AJ, BERGMANN S, BOCHUD M, BOERWINKLE E, BONNEFOND A, BONNYCASTLE LL, BORCH-JOHNSEN K, BOTTCHER Y, BRUNNER E, BUMPSTEAD SJ, CHARPENTIER G, CHEN YD, CHINES P, CLARKE R, COIN LJ, COOPER MN, CORNELIS M, CRAWFORD G, CRISPONI L, DAY IN, DE GEUS EJ, DELPLANQUE J, DINA C, ERDOS MR, FEDSON AC, FISCHER-ROSINSKY A, FOROUHI NG, FOX CS, FRANTS R, FRANZOSI MG, GALAN P, GOODARZI MO, GRAESSLER J, GROVES CJ, GRUNDY S, GWILLIAM R, GYLLENSTEN U, HADJADJ S, et al. 2010. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet, 42, 105–16. [PubMed: 20081858]

GALLAGHER EJ, LEROITH D, FRANCO R, ANTONIOU IM, NAYAK A, LIVAUDAIS-TOMAN J & BICKELL NA 2016. Metabolic syndrome and pre-diabetes contribute to racial disparities in breast cancer outcomes: hypothesis and proposed pathways. Diabetes Metab Res Rev, 32, 745–753. [PubMed: 26896340]

HU C & JIA W 2016. Linking MTNR1B Variants to Diabetes: The Role of Circadian Rhythms. Diabetes, 65, 1490–2. [PubMed: 27222395]

IARC 2007. Combined estrogen-progestogen contraceptives and combined estrogen-progestogen menopausal therapy. IARC Monogr Eval Carcinog Risks Hum, 91, 1–528. [PubMed: 18756632]

INUZUKA R, DILLER GP, BORGIA F, BENSON L, TAY EL, ALONSO-GONZALEZ R, SILVA M, CHARALAMBIDES M, SWAN L, DIMOPOULOS K & GATZOULIS MA 2012. Comprehensive use of cardiopulmonary exercise testing identifies adults with congenital heart disease at increased mortality risk in the medium term. Circulation, 125, 250–9. [PubMed: 22147905]

ISHWARAN H & KOGALUR UB 2007. Random Survival Forests for R Available: https://pdfs.semanticscholar.org/951a/84f0176076fb6786fdf43320e8b27094dcfa.pdf.

ISHWARAN H, KOGALUR UB, BLACKSTONE EH & LAUER MS 2008. RANDOM SURVIVAL FORESTS. 2 Available: https://projecteuclid.org/download/pdfview_1/euclid.aoas/1223908043.

JUNG SY, MANCUSO N, YU H, PAPP J, SOBEL E & ZHANG ZF 2019. Genome-Wide Meta-analysis of Gene-Environmental Interaction for Insulin Resistance Phenotypes and Breast Cancer Risk in Postmenopausal Women. Cancer Prev Res (Phila), 12, 31–42. [PubMed: 30327367]

KALLA SINGH S, BRITO C, TAN QW, DE LEON M & DE LEON D 2011. Differential expression and signaling activation of insulin receptor isoforms A and B: A link between breast cancer and diabetes. Growth Factors, 29, 278–89. [PubMed: 21913804]

KALLA SINGH S, TAN QW, BRITO C, DE LEON M & DE LEON D 2010. Insulin-like growth factors I and II receptors in the breast cancer survival disparity among African-American women. Growth Horm IGF Res, 20, 245–54. [PubMed: 20347606]

KEY TJ, APPLEBY PN, REEVES GK, RODDAM AW, HELZLSOUER KJ, ALBERG AJ, ROLLISON DE, DORGAN JF, BRINTON LA, OVERVAD K, KAAKS R, TRICHOPOULOU A, CLAVEL-CHAPELON F, PANICO S, DUELL EJ, PEETERS PH, RINALDI S, FENTIMAN IS, DOWSETT M, MANJER J, LENNER P, HALLMANS G, BAGLIETTO L, ENGLISH DR, GILES GG, HOPPER JL, SEVERI G, MORRIS HA, HANKINSON SE, TWOROGER SS, KOENIG K, ZELENIUCH-JACQUOTTE A, ARSLAN AA, TONIOLO P, SHORE RE, KROGH V, MICHELI A, BERRINO F, BARRETT-CONNOR E, LAUGHLIN GA, KABUTO M, AKIBA S, STEVENS RG, NERIISHI K, LAND CE, CAULEY JA, LUI LY, CUMMINGS SR, GUNTER MJ, ROHAN TE & STRICKLER HD 2011. Circulating sex hormones and breast cancer risk

factors in postmenopausal women: reanalysis of 13 studies. Br J Cancer, 105, 709–22. [PubMed: 21772329]

LAGOU V, MAGI R, HOTTENGA JJ, GRALLERT H, PERRY JRB, BOUATIA-NAJI N, MARULLO L, RYBIN D, JANSEN R, MIN JL, DIMAS AS, ULRICH A, ZUDINA L, GADIN JR, JIANG L, FAGGIAN A, BONNEFOND A, FADISTA J, STATHOPOULOU MG, ISAACS A, WILLEMS SM, NAVARRO P, TANAKA T, JACKSON AU, MONTASSER ME, O'CONNELL JR, BIELAK LF, WEBSTER RJ, SAXENA R, STAFFORD JM, POURCAIN BS, TIMPSON NJ, SALO P, SHIN SY, AMIN N, SMITH AV, LI G, VERWEIJ N, GOEL A, FORD I, JOHNSON PCD, JOHNSON T, KAPUR K, THORLEIFSSON G, STRAWBRIDGE RJ, RASMUSSEN-TORVIK LJ, ESKO T, MIHAILOV E, FALL T, FRASER RM, MAHAJAN A, KANONI S, GIEDRAITIS V, KLEBER ME, SILBERNAGEL G, MEYER J, MULLER-NURASYID M, GANNA A, SARIN AP, YENGO L, SHUNGIN D, LUAN J, HORIKOSHI M, AN P, SANNA S, BOETTCHER Y, RAYNER NW, NOLTE IM, ZEMUNIK T, IPEREN EV, KOVACS P, HASTIE ND, WILD SH, MCLACHLAN S, CAMPBELL S, POLASEK O, CARLSON O, EGAN J, KIESS W, WILLEMSEN G, KUUSISTO J, LAAKSO M, DIMITRIOU M, HICKS AA, RAURAMAA R, BANDINELLI S, THORAND B, LIU Y, MILJKOVIC I, LIND L, DONEY A, PEROLA M, HINGORANI A, KIVIMAKI M, KUMARI M, BENNETT AJ, GROVES CJ, HERDER C, KOISTINEN HA, KINNUNEN L, et al. 2021. Sex-dimorphic genetic effects and novel loci for fasting glucose and insulin variability. Nat Commun, 12, 24. [PubMed: 33402679]

LEON J, CASADO J, CARAZO A, SANJUAN L, MATE A, MUNOZ DE RUEDA P, DE LA CUEVA P, QUILES R, RUIZ S, RUIZ-EXTREMERA A & SALMERON J 2012. Gender-related invasion differences associated with mRNA expression levels of melatonin membrane receptors in colorectal cancer. Mol Carcinog, 51, 608–18. [PubMed: 21809392]

LIN R, YUAN Z, ZHANG C, JU H, SUN Y, HUANG N, CHEN L & JIN L 2020. Common genetic variants in ADCY5 and gestational glycemic traits. PLoS One, 15, e0230032. [PubMed: 32163478]

LIU L, WANG T, YANG X, XU C, LIAO Z, WANG X, SU D, LI Y, ZHOU H, QIU X, CHEN Y, HUANG D, LIAN C & SU P 2019. MTNR1B loss promotes chordoma recurrence by abrogating melatonin-mediated beta-catenin signaling repression. J Pineal Res, 67, e12588. [PubMed: 31140197]

MANNING AK, HIVERT MF, SCOTT RA, GRIMSBY JL, BOUATIA-NAJI N, CHEN H, RYBIN D, LIU CT, BIELAK LF, PROKOPENKO I, AMIN N, BARNES D, CADBY G, HOTTENGA JJ, INGELSSON E, JACKSON AU, JOHNSON T, KANONI S, LADENVALL C, LAGOU V, LAHTI J, LECOEUR C, LIU Y, MARTINEZ-LARRAD MT, MONTASSER ME, NAVARRO P, PERRY JR, RASMUSSEN-TORVIK LJ, SALO P, SATTAR N, SHUNGIN D, STRAWBRIDGE RJ, TANAKA T, VAN DUIJN CM, AN P, DE ANDRADE M, ANDREWS JS, ASPELUND T, ATALAY M, AULCHENKO Y, BALKAU B, BANDINELLI S, BECKMANN JS, BEILBY JP, BELLIS C, BERGMAN RN, BLANGERO J, BOBAN M, BOEHNKE M, BOERWINKLE E, BONNYCASTLE LL, BOOMSMA DI, BORECKI IB, BOTTCHER Y, BOUCHARD C, BRUNNER E, BUDIMIR D, CAMPBELL H, CARLSON O, CHINES PS, CLARKE R, COLLINS FS, CORBATON-ANCHUELO A, COUPER D, DE FAIRE U, DEDOUSSIS GV, DELOUKAS P, DIMITRIOU M, EGAN JM, EIRIKSDOTTIR G, ERDOS MR, ERIKSSON JG, EURY E, FERRUCCI L, FORD I, FOROUHI NG, FOX CS, FRANZOSI MG, FRANKS PW, FRAYLING TM, FROGUEL P, GALAN P, DE GEUS E, GIGANTE B, GLAZER NL, GOEL A, GROOP L, GUDNASON V, HALLMANS G, HAMSTEN A, HANSSON O, HARRIS TB, HAYWARD C, HEATH S, HERCBERG S, HICKS AA, HINGORANI A, HOFMAN A, HUI J, HUNG J, et al. 2012. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet, 44, 659–69. [PubMed: 22581228]

MCMULLAN CJ, SCHERNHAMMER ES, RIMM EB, HU FB & FORMAN JP 2013. Melatonin secretion and the incidence of type 2 diabetes. JAMA, 309, 1388–96. [PubMed: 23549584]

MOGENSEN UB, ISHWARAN H & GERDS TA 2012. Evaluating Random Forests for Survival Analysis using Prediction Error Curves. J Stat Softw, 50, 1–23. [PubMed: 25317082]

MONDAL AK, SHARMA NK, ELBEIN SC & DAS SK 2013. Allelic expression imbalance screening of genes in chromosome 1q21-24 region to identify functional variants for Type 2 diabetes susceptibility. Physiol Genomics, 45, 509–20. [PubMed: 23673729]

MONTAZERI M & BEIGZADEH A 2016. Machine learning models in breast cancer survival prediction. Technol Health Care, 24, 31–42. [PubMed: 26409558]

MUNSELL MF, SPRAGUE BL, BERRY DA, CHISHOLM G & TRENTHAM-DIETZ A 2014. Body mass index and breast cancer risk according to postmenopausal estrogen-progestin use and hormone receptor status. Epidemiol Rev, 36, 114–36. [PubMed: 24375928]

NATIONAL CANCER INSTITUTE June 1993. SEER Program: Comparative Staging Guide For Cancer

NISCHAN P, EBELING K, THOMAS DB & HIRSCH U 1993. Comparison of recalled and validated oral contraceptive histories. Am J Epidemiol, 138, 697–703. [PubMed: 8237985]

OBAYASHI S, HORIGUCHI J, HIGUCHI T, KATAYAMA A, HANDA T, ALTAN B, BAI T, BAO P, BAO H, YOKOBORI T, NISHIYAMA M, OYAMA T & KUWANO H 2017. Stathmin1 expression is associated with aggressive phenotypes and cancer stem cell marker expression in breast cancer patients. Int J Oncol, 51, 781–790. [PubMed: 28766688]

PANG H, LIN A, HOLFORD M, ENERSON BE, LU B, LAWTON MP, FLOYD E & ZHAO H 2006. Pathway analysis using random forests classification and regression. Bioinformatics, 22, 2028–36. [PubMed: 16809386]

PARK SL, CARMELLA SG, MING X, VIELGUTH E, STRAM DO, LE MARCHAND L & HECHT SS 2015. Variation in levels of the lung carcinogen NNAL and its glucuronides in the urine of cigarette smokers from five ethnic groups with differing risks for lung cancer. Cancer Epidemiol Biomarkers Prev, 24, 561–9. [PubMed: 25542827]

PARK SY, PALMER JR, ROSENBERG L, HAIMAN CA, BANDERA EV, BETHEA TN, TROESTER MA, VISCIDI E, KOLONEL LN, OLSHAN AF & AMBROSONE CB 2016. A case-control analysis of smoking and breast cancer in African American women: findings from the AMBER Consortium. Carcinogenesis, 37, 607–15. [PubMed: 27207658]

PATHANIA R, RAMACHANDRAN S, MARIAPPAN G, THAKUR P, SHI H, CHOI JH, MANICASSAMY S, KOLHE R, PRASAD PD, SHARMA S, LOKESHWAR BL, GANAPATHY V & THANGARAJU M 2016. Combined Inhibition of DNMT and HDAC Blocks the Tumorigenicity of Cancer Stem-like Cells and Attenuates Mammary Tumor Growth. Cancer Res, 76, 3224–35. [PubMed: 27197203]

RAMOS-MOLINA B, MARTIN MG & LINDBERG I 2016. PCSK1 Variants and Human Obesity. Prog Mol Biol Transl Sci, 140, 47–74. [PubMed: 27288825]

RAMOS E, CHEN G, SHRINER D, DOUMATEY A, GERRY NP, HERBERT A, HUANG H, ZHOU J, CHRISTMAN MF, ADEYEMO A & ROTIMI C 2011. Replication of genome-wide association studies (GWAS) loci for fasting plasma glucose in African-Americans. Diabetologia, 54, 783–8. [PubMed: 21188353]

ROBINSON WR, TSE CK, OLSHAN AF & TROESTER MA 2014. Body size across the life course and risk of premenopausal and postmenopausal breast cancer in Black women, the Carolina Breast Cancer Study, 1993–2001. Cancer Causes Control, 25, 1101–17. [PubMed: 24924530]

ROMAN TS, CANNON ME, VADLAMUDI S, BUCHKOVICH ML, WOLFORD BN, WELCH RP, MORKEN MA, KWON GJ, VARSHNEY A, KURSAWE R, WU Y, JACKSON AU, NATIONAL INSTITUTES OF HEALTH INTRAMURAL SEQUENCING CENTER COMPARATIVE SEQUENCING, P., ERDOS MR, KUUSISTO J, LAAKSO M, SCOTT LJ, BOEHNKE M, COLLINS FS, PARKER SCJ, STITZEL ML & MOHLKE KL 2017. A Type 2 Diabetes-Associated Functional Regulatory Variant in a Pancreatic Islet Enhancer at the ADCY5 Locus. Diabetes, 66, 2521–2530. [PubMed: 28684635]

ROSENBERG L, BETHEA TN, VISCIDI E, HONG CC, TROESTER MA, BANDERA EV, HAIMAN CA, KOLONEL LN, OLSHAN AF, AMBROSONE CB & PALMER JR 2016. Postmenopausal Female Hormone Use and Estrogen Receptor-Positive and -Negative Breast Cancer in African American Women. J Natl Cancer Inst, 108.

ROSENBERG MJ, LAYDE PM, ORY HW, STRAUSS LT, ROOKS JB & RUBIN GL 1983. Agreement between women's histories of oral contraceptive use and physician records. Int J Epidemiol, 12, 84–7. [PubMed: 6302013]

SAMSON ME, ADAMS SA, OREKOYA O & HEBERT JR 2016. Understanding the Association of Type 2 Diabetes Mellitus in Breast Cancer Among African American and European American Populations in South Carolina. J Racial Ethn Health Disparities, 3, 546–54. [PubMed: 27294746]

SCHUMACHER FR, AL OLAMA AA, BERNDT SI, BENLLOCH S, AHMED M, SAUNDERS EJ, DADAEV T, LEONGAMORNLERT D, ANOKIAN E, CIEZA-BORRELLA C, GOH C, BROOK MN, SHENG X, FACHAL L, DENNIS J, TYRER J, MUIR K, LOPHATANANON A, STEVENS VL, GAPSTUR SM, CARTER BD, TANGEN CM, GOODMAN PJ, THOMPSON IM JR., BATRA J, CHAMBERS S, MOYA L, CLEMENTS J, HORVATH L, TILLEY W, RISBRIDGER GP, GRONBERG H, ALY M, NORDSTROM T, PHAROAH P, PASHAYAN N, SCHLEUTKER J, TAMMELA TLJ, SIPEKY C, AUVINEN A, ALBANES D, WEINSTEIN S, WOLK A, HAKANSSON N, WEST CML, DUNNING AM, BURNET N, MUCCI LA, GIOVANNUCCI E, ANDRIOLE GL, CUSSENOT O, CANCEL-TASSIN G, KOUTROS S, BEANE FREEMAN LE, SORENSEN KD, ORNTOFT TF, BORRE M, MAEHLE L, GRINDEDAL EM, NEAL DE, DONOVAN JL, HAMDY FC, MARTIN RM, TRAVIS RC, KEY TJ, HAMILTON RJ, FLESHNER NE, FINELLI A, INGLES SA, STERN MC, ROSENSTEIN BS, KERNS SL, OSTRER H, LU YJ, ZHANG HW, FENG N, MAO X, GUO X, WANG G, SUN Z, GILES GG, SOUTHEY MC, MACINNIS RJ, FITZGERALD LM, KIBEL AS, DRAKE BF, VEGA A, GOMEZ-CAAMANO A, SZULKIN R, EKLUND M, KOGEVINAS M, LLORCA J, CASTANO-VINYALS G, PENNEY KL, STAMPFER M, PARK JY, SELLERS TA, LIN HY, STANFORD JL, CYBULSKI C, et al. 2018. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet, 50, 928–936. [PubMed: 29892016]

SCOTT RA, LAGOU V, WELCH RP, WHEELER E, MONTASSER ME, LUAN J, MAGI R, STRAWBRIDGE RJ, REHNBERG E, GUSTAFSSON S, KANONI S, RASMUSSEN-TORVIK LJ, YENGO L, LECOEUR C, SHUNGIN D, SANNA S, SIDORE C, JOHNSON PC, JUKEMA JW, JOHNSON T, MAHAJAN A, VERWEIJ N, THORLEIFSSON G, HOTTENGA JJ, SHAH S, SMITH AV, SENNBLAD B, GIEGER C, SALO P, PEROLA M, TIMPSON NJ, EVANS DM, POURCAIN BS, WU Y, ANDREWS JS, HUI J, BIELAK LF, ZHAO W, HORIKOSHI M, NAVARRO P, ISAACS A, O'CONNELL JR, STIRRUPS K, VITART V, HAYWARD C, ESKO T, MIHAILOV E, FRASER RM, FALL T, VOIGHT BF, RAYCHAUDHURI S, CHEN H, LINDGREN CM, MORRIS AP, RAYNER NW, ROBERTSON N, RYBIN D, LIU CT, BECKMANN JS, WILLEMS SM, CHINES PS, JACKSON AU, KANG HM, STRINGHAM HM, SONG K, TANAKA T, PEDEN JF, GOEL A, HICKS AA, AN P, MULLER-NURASYID M, FRANCO-CERECEDA A, FOLKERSEN L, MARULLO L, JANSEN H, OLDEHINKEL AJ, BRUINENBERG M, PANKOW JS, NORTH KE, FOROUHI NG, LOOS RJ, EDKINS S, VARGA TV, HALLMANS G, OKSA H, ANTONELLA M, NAGARAJA R, TROMPET S, FORD I, BAKKER SJ, KONG A, KUMARI M, GIGANTE B, HERDER C, MUNROE PB, CAULFIELD M, ANTTI J, MANGINO M, SMALL K, MILJKOVIC I, et al. 2012. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet, 44, 991–1005. [PubMed: 22885924]

SOULAKOVA JN, HARTMAN AM, LIU B, WILLIS GB & AUGUSTINE S 2012. Reliability of adult self-reported smoking history: data from the tobacco use supplement to the current population survey 2002–2003 cohort. Nicotine Tob Res, 14, 952–60. [PubMed: 22318688]

STEWART PA, LUKS J, ROYCIK MD, SANG QX & ZHANG J 2013. Differentially expressed transcripts and dysregulated signaling pathways and networks in African American breast cancer. PLoS One, 8, e82460. [PubMed: 24324792]

STRANGE KS, WILKINSON D, EDIN G & EMERMAN JT 2004. Mitogenic properties of insulin-like growth factors I and II, insulin-like growth factor binding protein-3 and epidermal growth factor on human breast stromal cells in primary culture. Breast Cancer Res Treat, 84, 77–84. [PubMed: 14999138]

STUMPF I, MUHLBAUER E & PESCHKE E 2008. Involvement of the cGMP pathway in mediating the insulin-inhibitory effect of melatonin in pancreatic beta-cells. J Pineal Res, 45, 318–27. [PubMed: 18363673]

SUGITA B, GILL M, MAHAJAN A, DUTTARGI A, KIROLIKAR S, ALMEIDA R, REGIS K, OLUWASANMI OL, MARCHI F, MARIAN C, MAKAMBI K, KALLAKURY B, SHEAHAN L, CAVALLI IJ, RIBEIRO EM, MADHAVAN S, BOCA S, GUSEV Y & CAVALLI LR 2016. Differentially expressed miRNAs in triple negative breast cancer between African-American and non-Hispanic white women. Oncotarget, 7, 79274–79291. [PubMed: 27813494]

TERRY PD, MILLER AB & ROHAN TE 2002. Cigarette smoking and breast cancer risk: a long latency period? Int J Cancer, 100, 723–8. [PubMed: 12209614]

TONG X, FENG Y & LI JJ 2018. Neyman-Pearson classification algorithms and NP receiver operating characteristics. Sci Adv, 4, eaao1659. [PubMed: 29423442]

TRENTHAM-DIETZ A, NEWCOMB PA, STORER BE & REMINGTON PL 2000. Risk factors for carcinoma in situ of the breast. Cancer Epidemiol Biomarkers Prev, 9, 697–703. [PubMed: 10919740]

URBAN M, BANKS E, EGGER S, CANFELL K, O'CONNELL D, BERAL V & SITAS F 2012. Injectable and oral contraceptive use and cancers of the breast, cervix, ovary, and endometrium in black South African women: case-control study. PLoS Med, 9, e1001182. [PubMed: 22412354]

VALLEGA KA, LIU N, MYERS JS, YU K & SANG QX 2016. Elevated Resistin Gene Expression in African American Estrogen and Progesterone Receptor Negative Breast Cancer. PLoS One, 11, e0157741. [PubMed: 27314854]

WANG Q, ZHU Y, LI Z, BU Q, SUN T, WANG H, SUN H & CAO X 2019. Up-regulation of SPC25 promotes breast cancer. Aging (Albany NY), 11, 5689–5704. [PubMed: 31400751]

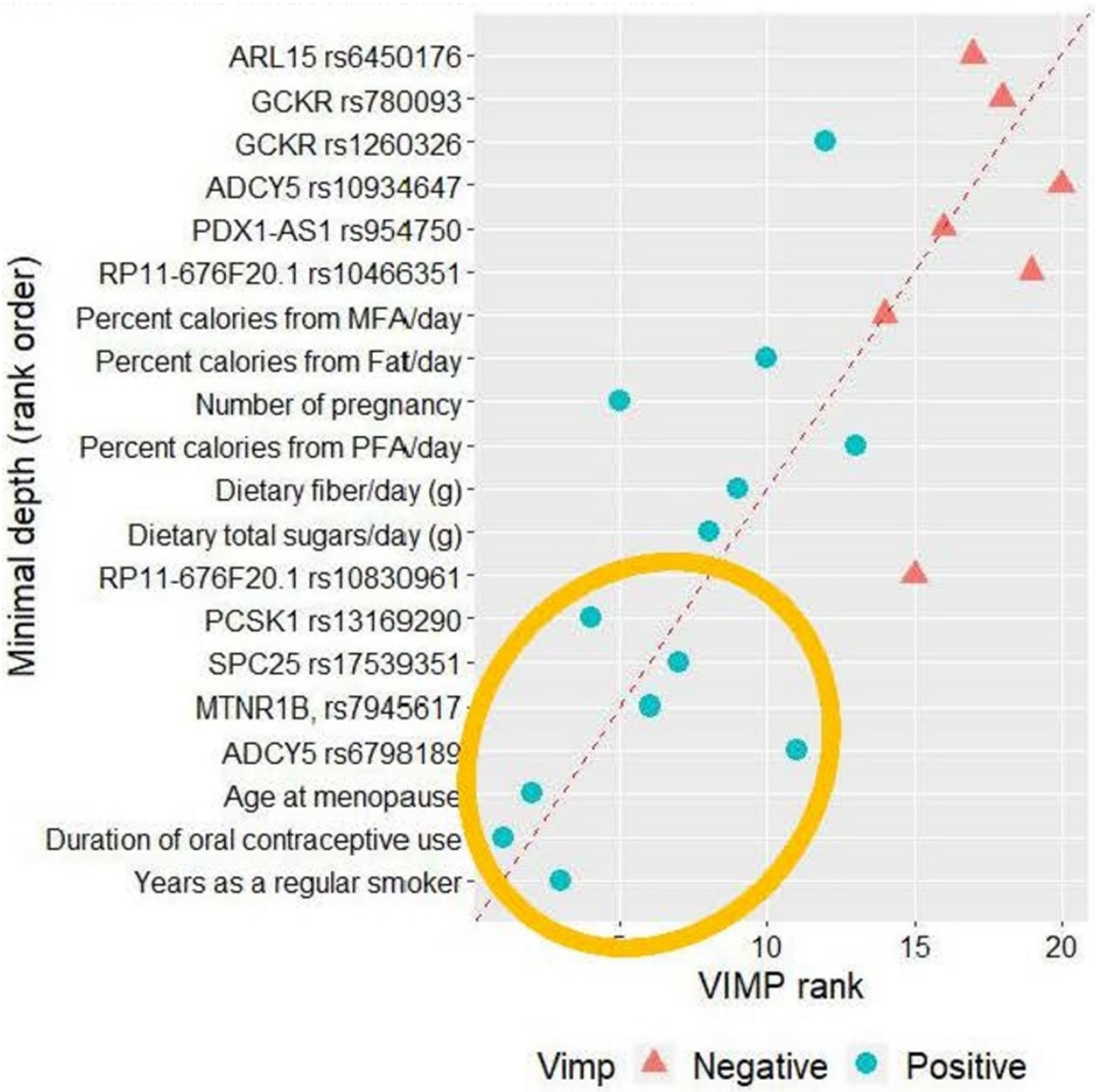
WHITE KK, PARK SY, KOLONEL LN, HENDERSON BE & WILKENS LR 2012. Body size and breast cancer risk: the Multiethnic Cohort. Int J Cancer, 131, E705–16. [PubMed: 22120517]

WILLIAMS LA, OLSHAN AF, HONG CC, BANDERA EV, ROSENBERG L, CHENG TD, LUNETTA KL, MCCANN SE, POOLE C, KOLONEL LN, PALMER JR, AMBROSONE CB & TROESTER MA 2017. Alcohol Intake and Breast Cancer Risk in African American Women from the AMBER Consortium. Cancer Epidemiol Biomarkers Prev, 26, 787–794. [PubMed: 28420652]

WILLIAMS LA, OLSHAN AF, TSE CK, BELL ME & TROESTER MA 2016. Alcohol intake and invasive breast cancer risk by molecular subtype and race in the Carolina Breast Cancer Study. Cancer Causes Control, 27, 259–69. [PubMed: 26705260]

XING CY, DOOSE M, QIN B, LIN Y, PLASCAK JJ, OMENE C, HE C, DEMISSIE K, HONG CC, BANDERA EV & LLANOS AAM 2020. Prediagnostic Allostatic Load as a Predictor of Poorly Differentiated and Larger Sized Breast Cancers among Black Women in the Women's Circle of Health Follow-Up Study. Cancer Epidemiol Biomarkers Prev, 29, 216–224. [PubMed: 31719063]

YUAN L, COLLINS AR, DAI J, DUBOCOVICH ML & HILL SM 2002. MT(1) melatonin receptor overexpression enhances the growth suppressive effect of melatonin in human breast cancer cells. Mol Cell Endocrinol, 192, 147–56. [PubMed: 12088876]



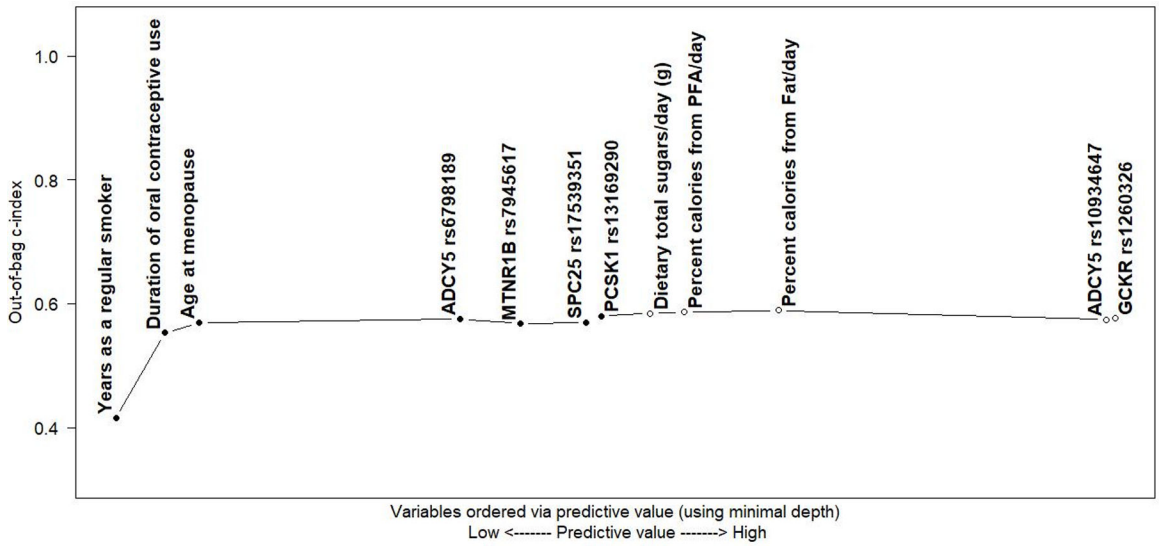
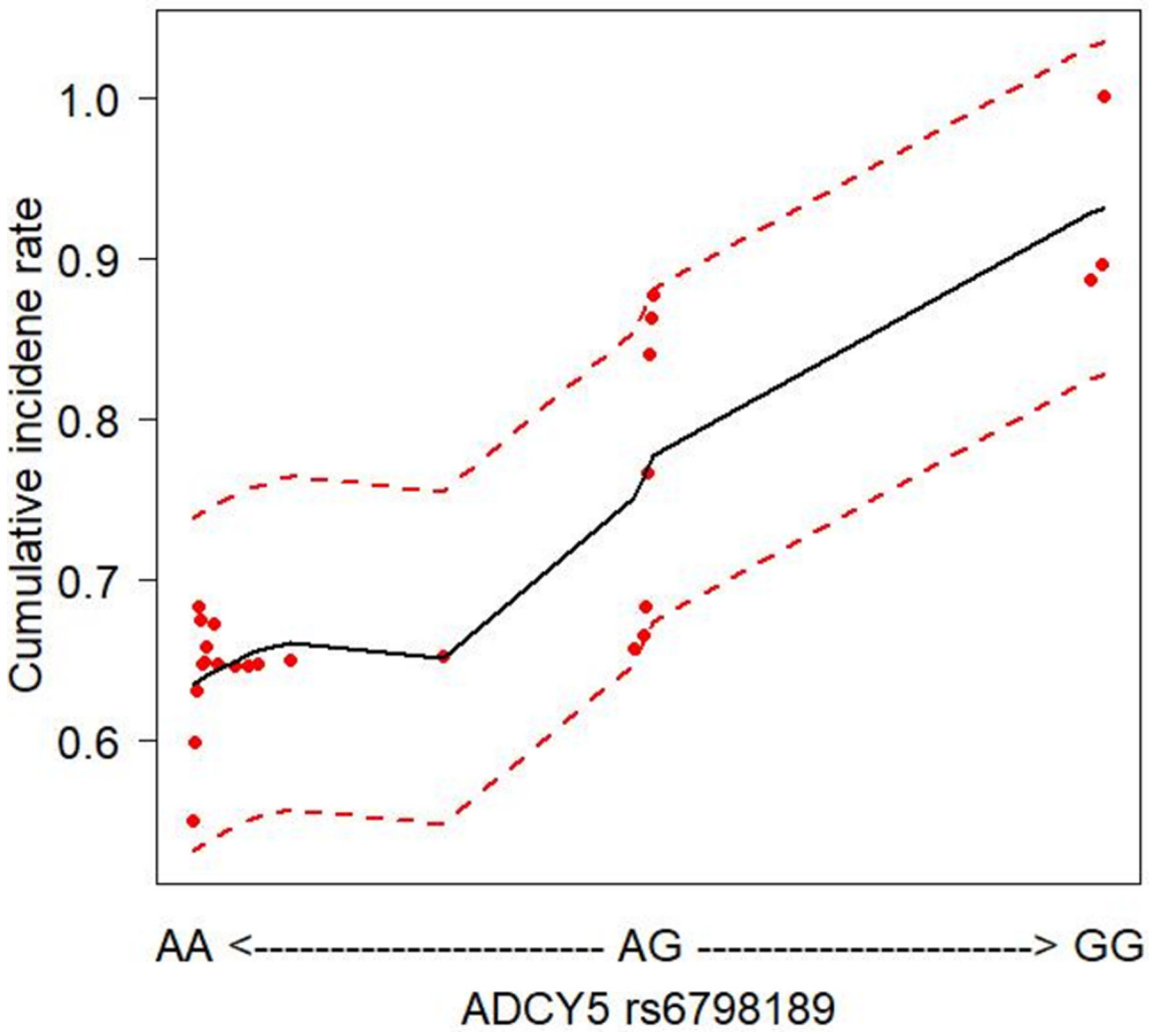


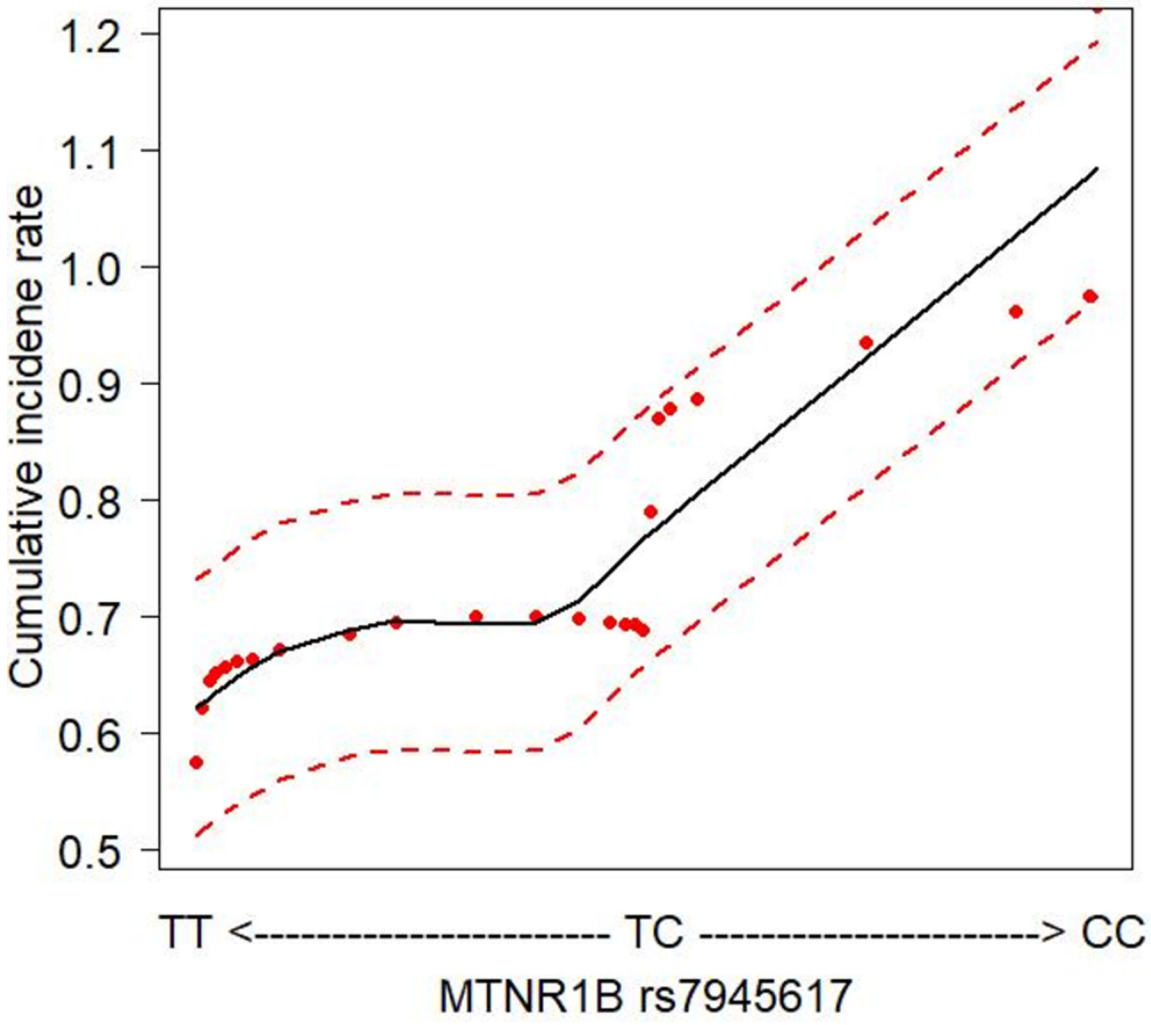
Figure 1.

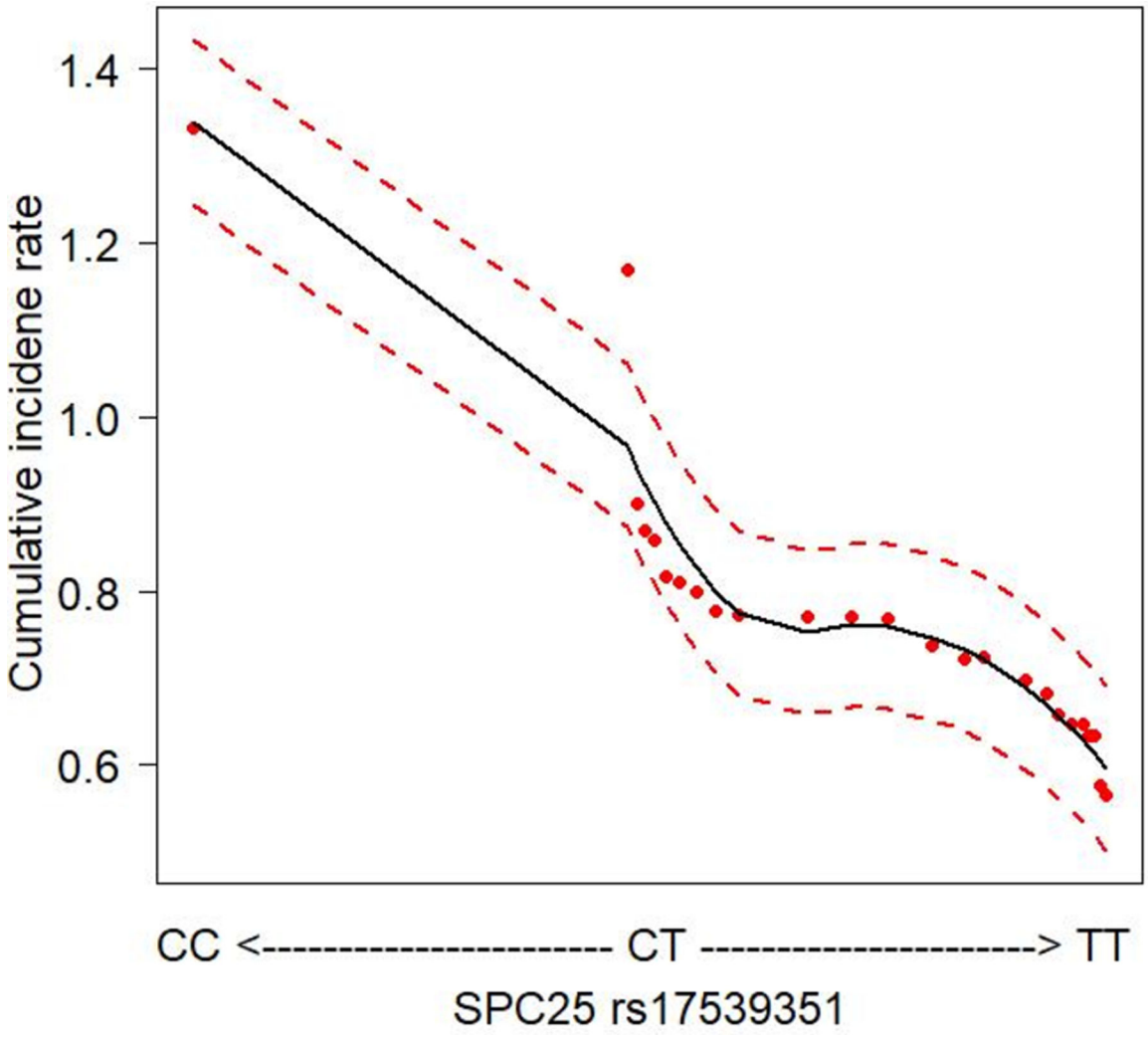
The second stage of random survival forest (RSF) using 11 single-nucleotide polymorphisms and 9 behavioral factors selected from the first stage of RSF analysis

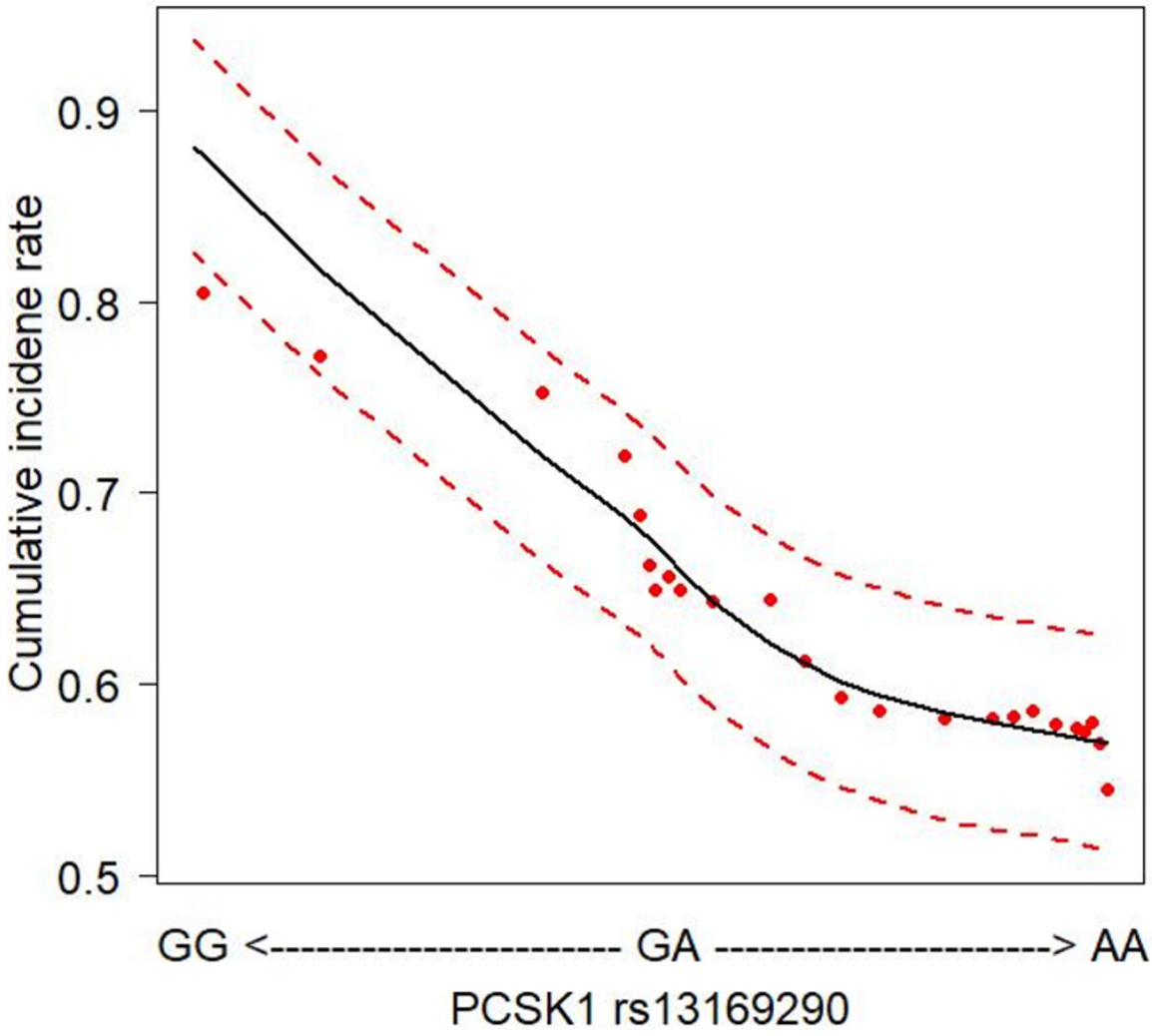
A. Comparison of minimal depth and VIMP rankings. (MFA, monounsaturated fatty acid; PFA, polyunsaturated fatty acid; VIMP, variable of importance. Note: The 7 variables within the gold ellipse were identified as the most influential predictors.)

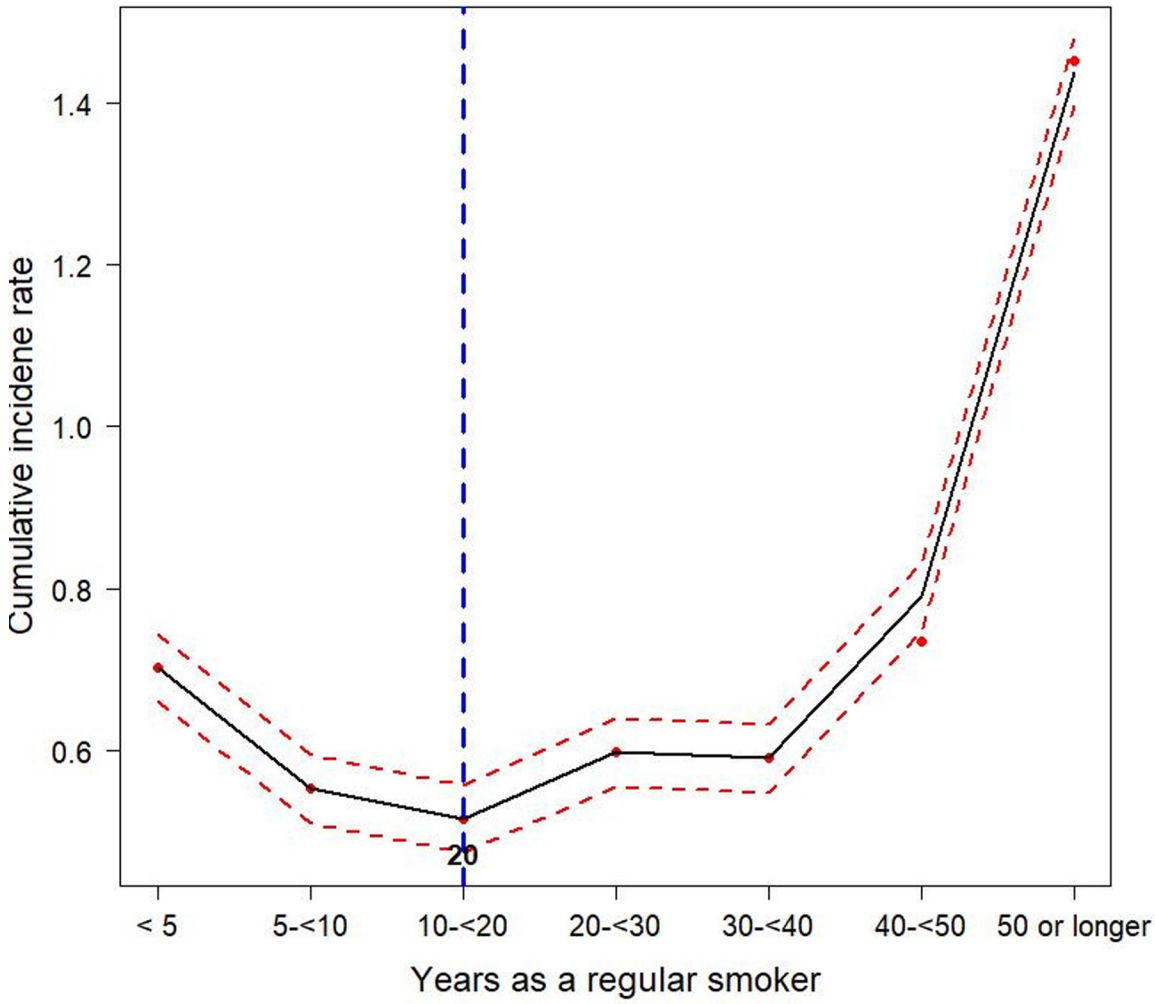
B. Out-of-bag concordance index (c-index). (An improvement in c-index was observed when the top 7 variables [●] were added to the model, whereas other variables [○] did not further improve the accuracy of prediction.)

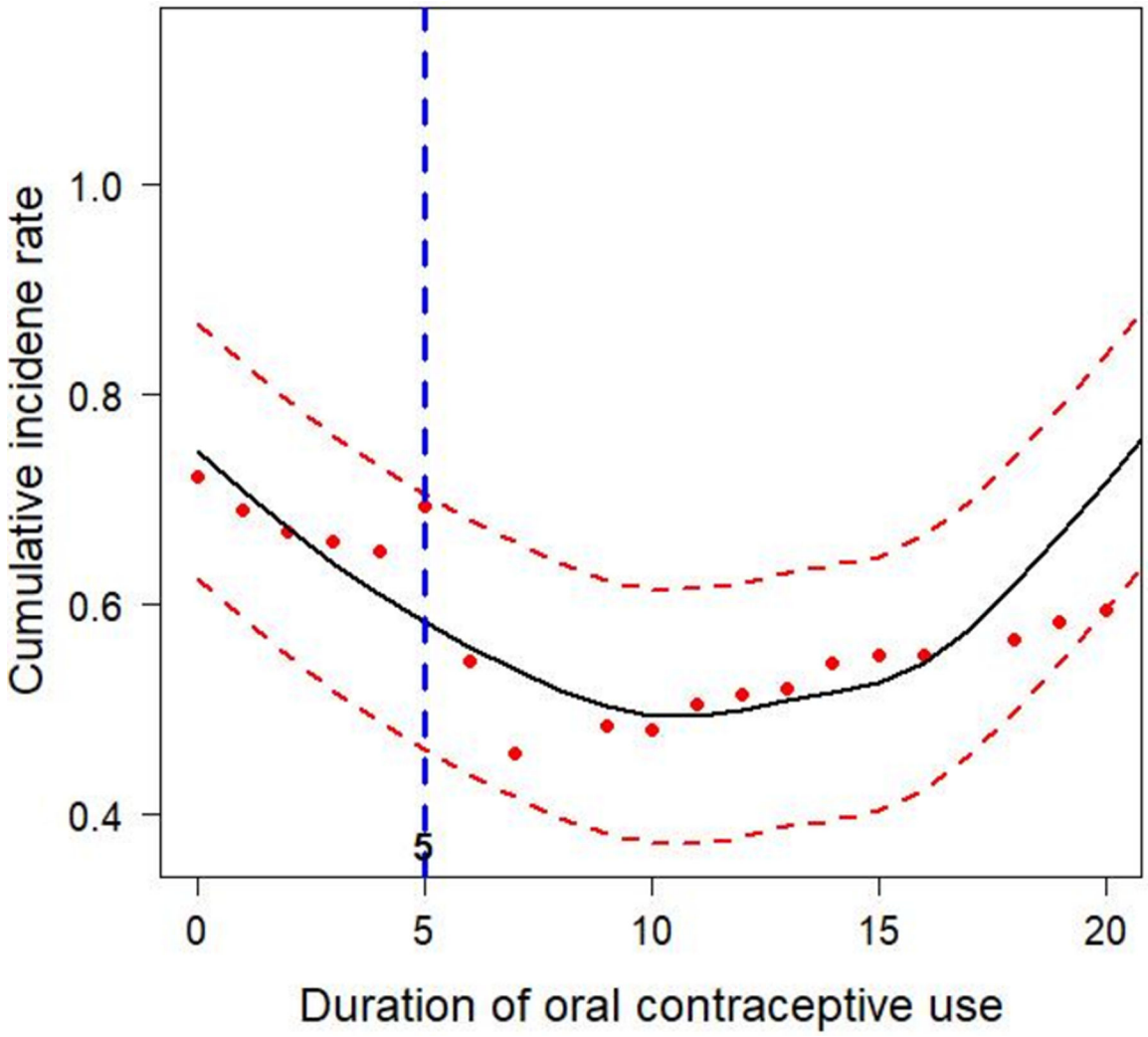












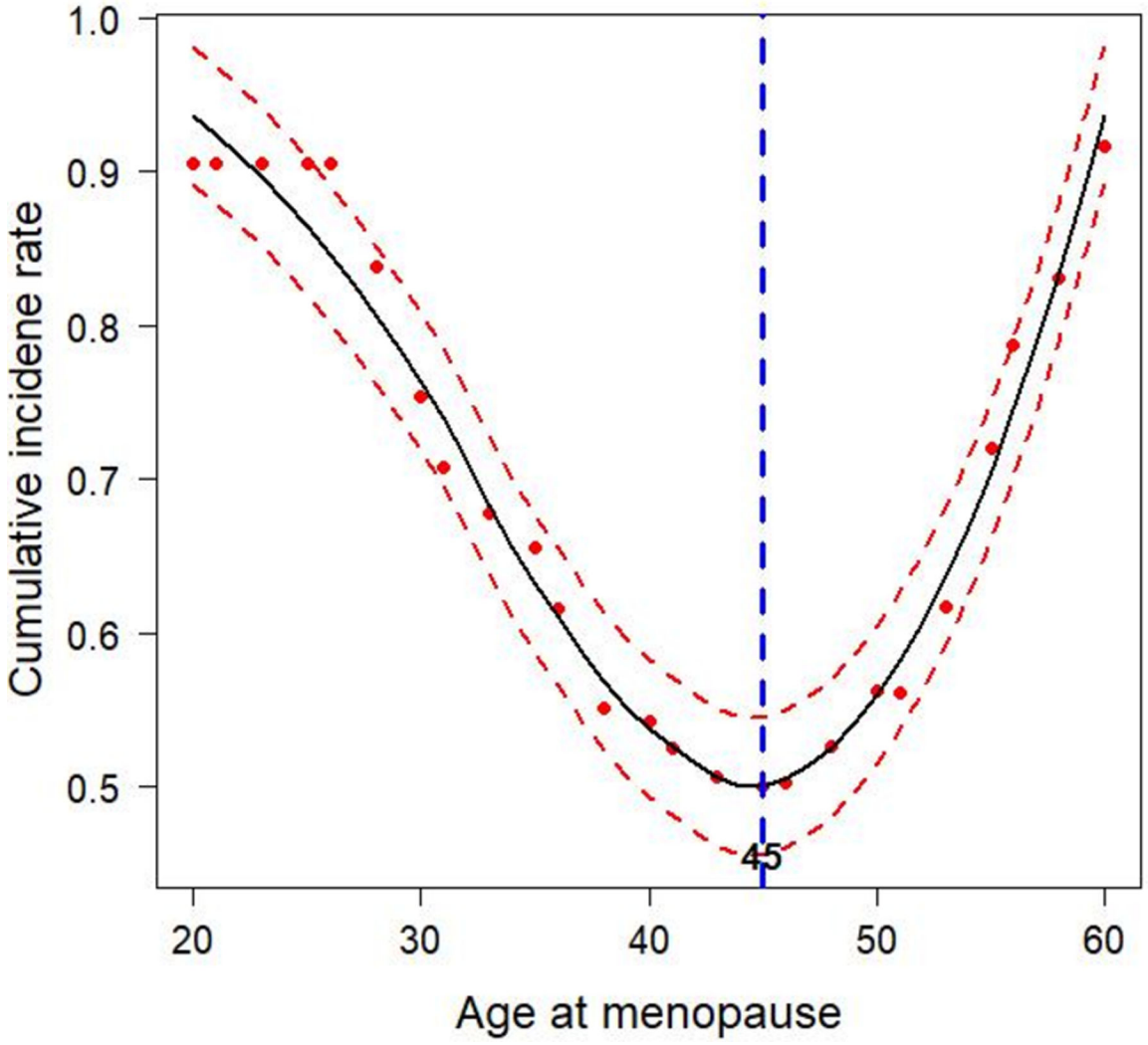


Figure 2.
Cumulative breast cancer incidence rate for the 7 most predictive variables (4 single-nucleotide polymorphisms and 3 behavioral factors) based on a random survival forest analysis. (Dashed red lines indicate 95% confidence intervals.)

Table 1.

The second stage of RSF analysis: predictive values of variables for breast cancer risk

Variable*	Minimal Depth†	VIMP	C-index	Incremental Error¶	Drop Error§
Years as a regular smoker	**2.7101**	**0.0063**	**0.4155**	**0.5845**	**−0.0845**
Duration of oral contraceptive use	**2.8196**	**0.0106**	**0.5526**	**0.4474**	**0.1371**
Age at menopause	**2.8980**	**0.0070**	**0.5694**	**0.4306**	**0.0168**
***ADCY5* rs6798189**	**3.4848**	**0.0008**	**0.5753**	**0.4247**	**0.0059**
***MTNR1B* rs7945617**	**3.6198**	**0.0026**	**0.5675**	**0.4325**	**−0.0077**
***SPC25* rs17539351**	**3.7691**	**0.0014**	**0.5698**	**0.4302**	**0.0022**
***PCSK1* rs13169290**	**3.8024**	**0.0044**	**0.5798**	**0.4202**	**0.0100**
RP11-676F20.1 rs10830961	3.8360	−0.0001	0.5806	0.4194	0.0009
Dietary total sugars/day (g)	3.9134	0.0013	0.5841	0.4159	0.0035
Dietary fiber/day (g)	3.9811	0.0011	0.5769	0.4231	−0.0073
Percent calories from PFA/day	3.9894	0.0005	0.5865	0.4135	0.0096
Number of pregnancies	4.1099	0.0043	0.5970	0.4030	0.0106
Percent calories from fat/day	4.2016	0.0009	0.5891	0.4109	−0.0080
Percent calories from MFA/day	4.2053	0.0000	0.5871	0.4129	−0.0020
RP11-676F20.1 rs10466351	4.2344	−0.0019	0.5819	0.4181	−0.0052
PDX1-AS1 rs954750	4.2351	−0.0003	0.5836	0.4164	0.0017
ADCY5 rs10934647	4.9397	−0.0021	0.5740	0.4260	−0.0096
GCKR rs1260326	4.9595	0.0007	0.5763	0.4237	0.0023
GCKR rs780093	5.6440	−0.0008	0.5681	0.4319	−0.0082
ARL15 rs6450176	5.8423	−0.0007	0.5678	0.4322	−0.0003

C-index, concordance index; MFA, monounsaturated fatty acid; PFA, polyunsaturated fatty acid; RSF, random survival forest; VIMP, variable of importance. Variables in bold face were selected as the most predictive markers on the basis of multimodal predictive values.

*Variables ordered by minimal depth.

†Minimal depth estimated as the predictive value of the variable in the nested RSF models, with a lower value likely to have a greater impact on prediction.

¶The incremental error rate was computed in the nested sequence of models starting with the top variable, followed by the model with the top 2 variables, then the model with the top 3 variables, and so on. For example, the third error rate was computed from the third nested model (including the first, second, and third variables).

§The drop error rate of the variable was computed as the difference between the error rates of a prior and the corresponding variable from the nested models. For example, the drop error rate of the second variable was estimated by the difference between the error rates from the first and second nested models. The error rate for the null model is set at 0.5; thus, the drop error rate for the first variable was obtained by subtracting the error rate (0.5845) from 0.5.

Table 2.

Women experiencing menopause at ≤ 45 years: combined and joint effects of smoking with risk genotypes and risk behaviors predicting breast cancer risk

	Age at menopause of ≤ 45 years		**Never Smokers**			**Regular smokers for ≥ 20 years**		
n of risks	**HR[†] (95% CI)**	***p***	**n**	**HR[†] (95% CI)**	***p***	**n**	**HR[†] (95% CI)**	***p***
Risk genotypes[£]								
0	reference		210	reference		500	2.28 (0.78 - 6.73)	0.134
1	1.17 (0.71 - 1.93)	0.537	168	1.24 (0.31 - 4.97)	0.761	380	**3.22 (1.10 - 9.47)**	**0.033**
2+	**2.05 (1.08 - 3.90)**	**0.029**	44	2.39 (0.44 - 13.07)	0.314	99	**4.82 (1.45 - 16.09)**	**0.010**
p_{trend}		0.400						
Risk behaviors[¶]								
0	reference		242	reference		549	2.48 (0.94 - 6.54)	0.066
1	1.72 (0.91 - 3.22)	0.092	180	1.39 (0.40 - 4.81)	0.604	428	**3.03 (1.15 – 8.00)**	**0.025**
2	**2.23 (1.13 - 4.43)**	**0.021**[*]						
p_{trend}		0.300						
Risk genotypes combined with behavioral factors[§]								
0	reference		129	reference		279	0.93 (0.27 - 3.20)	0.910
1	1.12 (0.51 - 2.45)	0.771	194	0.17 (0.02 - 1.48)	0.108	489	2.22 (0.78 - 6.36)	0.136
2	1.95 (0.91 - 4.20)	0.087	99	1.67 (0.45 - 6.25)	0.444	209	1.96 (0.62 - 6.20)	0.250
p_{trend}		0.300						

CI, confidence interval; FDR, false discovery rate; HR, hazard ratio. Numbers in bold face are statistically significant.

[†]Multivariate regression for risk genotype and behaviors was adjusted by dietary fiber/day (g), dietary total sugars/day (g), percent calories from polyunsaturated fatty acid/day, percent calories from monounsaturated fatty acid/day, and number of pregnancies.

[*] *p* values with FDR < 0.05 were presented after being corrected for multiple testing via the Benjamini-Hochberg method.

[£]The number of risk genotypes (*ADCY5* rs6798189 AG+GG; *MTNR1B* rs7945617 CC; *SPC25* rs17539351 TC+CC; and *PCSK1* rs13169290 AG+GG) was defined as follows: 0 (none) vs. 1 (1 risk allele) vs. 2+ (2 or more risk alleles).

[¶]The number of behaviors (years as a regular smoker < 20 years vs. ≥ 20 years [in overall analysis only] and duration of oral contraceptive use 5 >– 20 years vs. ≤ 5 years) was defined as follows: 0 (null risk behavior) vs. 1 (1 risk behavior) vs. 2 (2 risk behaviors).

§The combined number of risk genotypes and risk behaviors was based on risk genotypes defined as 0 (none) and 1+ (1 or more risk alleles) and based on risk behaviors defined as 0 (none) and 1+ (1 or more risk behaviors). The ultimate number of risk genotypes combined with risk behaviors was defined as 0 (none of risk genotypes and behaviors), 1 (either risk genotypes or risk behaviors), and 2 (both risk genotypes and behaviors).

Table 3.

Women experiencing menopause at ≤ 45 years: combined and joint effects of smoking with risk genotypes and risk behaviors predicting breast cancer risk

	Age at menopause of ≤ 45 years		**Regular smokers for < 20 years**			**Regular smokers for ≥ 20 years**		
n of risks	**HR[†] (95% CI)**	***p***	**n**	**HR[†] (95% CI)**	***p***	**n**	**HR[†] (95% CI)**	***p***
Risk genotypes[£]								
0	reference		227	reference		500	0.94 (0.44 - 2.03)	0.882
1	1.17 (0.71 - 1.93)	0.537	151	0.74 (0.25 - 2.16)	0.577	380	1.32 (0.62 - 2.83)	0.473
2+	**2.05 (1.08 - 3.90)**	**0.029**	36	1.94 (0.53 - 7.04)	0.317	99	2.01 (0.79 - 5.12)	0.143
p_{trend}		0.400						
Risk behaviors[¶]								
0	reference		231	reference		549	1.40 (0.63 - 3.13)	0.407
1	1.72 (0.91 - 3.22)	0.092	182	1.63 (0.64 - 4.14)	0.303	428	1.75 (0.78 - 3.92)	0.176
2	**2.23 (1.13 - 4.43)**	**0.021**[*]						
p_{trend}		0.300						
Risk genotypes combined with behavioral factors[§]								
0	reference		125	reference		279	0.85 (0.25 - 2.92)	0.802
1	1.12 (0.51 - 2.45)	0.771	207	1.52 (0.47 - 4.84)	0.482	489	2.09 (0.73 - 5.96)	0.168
2	1.95 (0.91 - 4.20)	0.087	81	1.57 (0.39 - 6.31)	0.521	209	1.83 (0.58 - 5.75)	0.304
p_{trend}		0.300						

CI, confidence interval; FDR, false discovery rate; HR, hazard ratio. Numbers in bold face are statistically significant.

[†]Multivariate regression for risk genotype and behaviors was adjusted by dietary fiber/day (g), dietary total sugars/day (g), percent calories from polyunsaturated fatty acid/day, percent calories from monounsaturated fatty acid/day, and number of pregnancies.

[*] *p* values with FDR < 0.05 were presented after being corrected for multiple testing via the Benjamini-Hochberg method.

[£]The number of risk genotypes (*ADCY5* rs6798189 AG+GG; *MTNR1B* rs7945617 CC; *SPC25* rs17539351 TC+CC; and *PCSK1* rs13169290 AG+GG) was defined as follows: 0 (none) vs. 1 (1 risk allele) vs. 2+ (2 or more risk alleles).

[¶]The number of behaviors (years as a regular smoker < 20 years vs. ≥ 20 years [in overall analysis only] and duration of oral contraceptive use 5 >– 20 years vs. ≤ 5 years) was defined as follows: 0 (null risk behavior) vs. 1 (1 risk behavior) vs. 2 (2 risk behaviors).

§The combined number of risk genotypes and risk behaviors was based on risk genotypes defined as 0 (none) and 1+ (1 or more risk alleles) and based on risk behaviors defined as 0 (none) and 1+ (1 or more risk behaviors). The ultimate number of risk genotypes combined with risk behaviors was defined as 0 (none of risk genotypes and behaviors), 1 (either risk genotypes or risk behaviors), and 2 (both risk genotypes and behaviors).

Table 4.

Women experiencing menopause at > 45 years: combined and joint effects of smoking with risk genotypes and risk behaviors predicting breast cancer risk

	Age at menopause of > 45 years		**Never Smokers**			**Regular smokers for ≥ 20 years**		
n of risks	**HR[†] (95% CI)**	***p***	**n**	**HR[†] (95% CI)**	***p***	**n**	**HR[†] (95% CI)**	***p***
Risk genotypes[£]								
0	reference		329	reference		748	1.97 (0.90 - 4.30)	0.089
1	**1.53 (1.06 - 2.21)**	**0.022** *	302	2.06 (0.87 - 4.87)	0.098	628	**2.69 (1.25 - 5.82)**	**0.012** *
2+	**2.14 (1.33 - 3.45)**	**0.002** *	79	**3.66 (1.32 - 10.11)**	**0.012** *	171	**4.02 (1.70 - 9.50)**	**1.51E-03** *
p_{trend}		0.100						
Risk behaviors[¶]								
0	reference		433	reference		873	1.63 (0.84 - 3.15)	0.148
1	**1.84 (1.15 - 2.95)**	**0.011** *	275	**2.40 (1.15 - 4.99)**	**0.019** *	669	**2.89 (1.53 - 5.48)**	**1.14E-03** *
2	**2.63 (1.59 - 4.36)**	**1.70E-04** *						
p_{trend}		**0.030**						
Risk genotypes combined with behavioral factors[§]								
0	reference		214	reference		420	2.62 (0.56 - 12.14)	0.219
1	**2.47 (1.13 - 5.41)**	**0.024** *	333	**5.21 (1.20 - 22.68)**	**0.028** *	778	**7.54 (1.83 - 31.09)**	**0.005** *
2	**4.01 (1.85 - 8.71)**	**4.51E-04** *	161	**8.19 (1.83 - 36.66)**	**0.006** *	344	**8.67 (2.04 - 36.75)**	**0.003** *
p_{trend}		**0.010**						

CI, confidence interval; FDR, false discovery rate; HR, hazard ratio. Numbers in bold face are statistically significant.

[†]Multivariate regression for risk genotype and behaviors was adjusted by dietary fiber/day (g), dietary total sugars/day (g), percent calories from polyunsaturated fatty acid/day, percent calories from monounsaturated fatty acid/day, and number of pregnancies.

[*]*p* values with FDR < 0.05 were presented after being corrected for multiple testing via the Benjamini-Hochberg method.

[£]The number of risk genotypes (*ADCY5* rs6798189 AG+GG; *MTNR1B* rs7945617 CC; *SPC25* rs17539351 TC+CC; and *PCSK1* rs13169290 AG+GG) was defined as follows: 0 (none) vs. 1 (1 risk allele) vs. 2+ (2 or more risk alleles).

[¶]The number of behaviors (years as a regular smoker < 20 years vs. ≥ 20 years [in overall analysis only] and duration of oral contraceptive use 5 >– 20 years vs. ≤ 5 years) was defined as follows: 0 (null risk behavior) vs. 1 (1 risk behavior) vs. 2 (2 risk behaviors).

[§]The combined number of risk genotypes and risk behaviors was based on risk genotypes defined as 0 (none) and 1+ (1 or more risk alleles) and based on risk behaviors defined as 0 (none) and 1+ (1 or more risk behaviors). The ultimate number of risk genotypes combined with risk behaviors was defined as 0 (none of risk genotypes and behaviors), 1 (either risk genotypes or risk behaviors), and 2 (both risk genotypes and behaviors).

Table 5.

Women experiencing menopause at > 45 years: combined and joint effects of smoking with risk genotypes and risk behaviors predicting breast cancer risk

	Age at menopause of > 45 years		**Regular smokers for < 20 years**			**Regular smokers for ≥ 20 years**		
n of risks	**HR[†] (95% CI)**	***p***	**n**	**HR[†] (95% CI)**	***p***	**n**	**HR[†] (95% CI)**	***p***
Risk genotypes[£]								
0	reference		314	reference		748	1.13 (0.58 - 2.21)	0.711
1	**1.53 (1.06 - 2.21)**	**0.022** *	237	1.63 (0.76 - 3.48)	0.209	628	1.57 (0.81 - 3.01)	0.179
2+	**2.14 (1.33 - 3.45)**	**0.002** *	66	1.20 (0.34 - 4.24)	0.782	171	**2.37 (1.11 - 5.07)**	**0.026**
p_{trend}		0.100						
Risk behaviors[¶]								
0	reference		349	reference		873	1.33 (0.68 - 2.63)	0.407
1	**1.84 (1.15 - 2.95)**	**0.011** *	265	**2.36 (1.12 - 4.96)**	**0.024** *	669	**2.39 (1.23 - 4.63)**	**0.010** *
2	**2.63 (1.59 - 4.36)**	**1.70E-04** *						
p_{trend}		**0.030**						
Risk genotypes combined with behavioral factors[§]								
0	reference		167	reference		420	0.75 (0.25 - 2.24)	0.606
1	**2.47 (1.13 - 5.41)**	**0.024** *	328	1.35 (0.48 - 3.79)	0.568	778	2.19 (0.87 - 5.50)	0.096
2	**4.01 (1.85 - 8.71)**	**4.51E-04** *	119	**3.42 (1.20 - 9.73)**	**0.021** *	344	2.53 (0.96 - 6.65)	0.060
p_{trend}		**0.010**						

CI, confidence interval; FDR, false discovery rate; HR, hazard ratio. Numbers in bold face are statistically significant.

[†]Multivariate regression for risk genotype and behaviors was adjusted by dietary fiber/day (g), dietary total sugars/day (g), percent calories from polyunsaturated fatty acid/day, percent calories from monounsaturated fatty acid/day, and number of pregnancies.

[*]*p* values with FDR < 0.05 were presented after being corrected for multiple testing via the Benjamini-Hochberg method.

[£]The number of risk genotypes (*ADCY5* rs6798189 AG+GG; *MTNR1B* rs7945617 CC; *SPC25* rs17539351 TC+CC; and *PCSK1* rs13169290 AG+GG) was defined as follows: 0 (none) vs. 1 (1 risk allele) vs. 2+ (2 or more risk alleles).

[¶]The number of behaviors (years as a regular smoker < 20 years vs. ≥ 20 years [in overall analysis only] and duration of oral contraceptive use 5 >– 20 years vs. ≤ 5 years) was defined as follows: 0 (null risk behavior) vs. 1 (1 risk behavior) vs. 2 (2 risk behaviors).

[§]The combined number of risk genotypes and risk behaviors was based on risk genotypes defined as 0 (none) and 1+ (1 or more risk alleles) and based on risk behaviors defined as 0 (none) and 1+ (1 or more risk behaviors). The ultimate number of risk genotypes combined with risk behaviors was defined as 0 (none of risk genotypes and behaviors), 1 (either risk genotypes or risk behaviors), and 2 (both risk genotypes and behaviors).