



# Synergistic Effects of Genetic Variants of Glucose Homeostasis and Lifelong Exposures to Cigarette Smoking, Female Hormones, and Dietary Fat Intake on Primary Colorectal Cancer Development in African and Hispanic/Latino American Women

OPEN ACCESS

**Edited by:**

Farnam Mohebi,  
University of California, Berkeley,  
United States

**Reviewed by:**

Athena Starlard-Davenport,  
University of Tennessee Health  
Science Center (UTHSC),  
United States  
Mujeeb Zafar Banday,  
Government Medical College (GMC),  
India

**\*Correspondence:**

Su Yon Jung  
sjung@sonnet.ucla.edu

**Specialty section:**

This article was submitted to  
Cancer Epidemiology and Prevention,  
a section of the journal  
Frontiers in Oncology

**Received:** 17 August 2021

**Accepted:** 22 September 2021

**Published:** 07 October 2021

**Citation:**

Jung SY, Sobel EM, Pellegrini M, Yu H  
and Papp JC (2021) Synergistic  
Effects of Genetic Variants of Glucose  
Homeostasis and Lifelong Exposures  
to Cigarette Smoking, Female  
Hormones, and Dietary Fat Intake on  
Primary Colorectal Cancer  
Development in African and Hispanic/  
Latino American Women.  
*Front. Oncol.* 11:760243.  
doi: 10.3389/fonc.2021.760243

Su Yon Jung<sup>1\*</sup>, Eric M. Sobel<sup>2,3</sup>, Matteo Pellegrini<sup>4</sup>, Herbert Yu<sup>5</sup> and Jeanette C. Papp<sup>2</sup>

<sup>1</sup> Translational Sciences Section, Jonsson Comprehensive Cancer Center, School of Nursing, University of California, Los Angeles, Los Angeles, CA, United States, <sup>2</sup> Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States, <sup>3</sup> Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States, <sup>4</sup> Department of Molecular, Cell and Developmental Biology, Life Sciences Division, University of California, Los Angeles, Los Angeles, CA, United States, <sup>5</sup> Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, United States

**Background:** Disparities in cancer genomic science exist among racial/ethnic minorities. Particularly, African American (AA) and Hispanic/Latino American (HA) women, the 2 largest minorities, are underrepresented in genetic/genome-wide studies for cancers and their risk factors. We conducted on AA and HA postmenopausal women a genomic study for insulin resistance (IR), the main biologic mechanism underlying colorectal cancer (CRC) carcinogenesis owing to obesity.

**Methods:** With 780 genome-wide IR-specific single-nucleotide polymorphisms (SNPs) among 4,692 AA and 1,986 HA women, we constructed a CRC-risk prediction model. Along with these SNPs, we incorporated CRC-associated lifestyles in the model of each group and detected the topmost influential genetic and lifestyle factors. Further, we estimated the attributable risk of the topmost risk factors shared by the groups to explore potential factors that differentiate CRC risk between these groups.

**Results:** In both groups, we detected IR-SNPs in *PCSK1* (in AA) and *IFT172*, *GCKR*, and *NRBP1* (in HA) and risk lifestyles, including long lifetime exposures to cigarette smoking and endogenous female hormones and daily intake of polyunsaturated fatty acids (PFA), as the topmost predictive variables for CRC risk. Combinations of those top genetic- and lifestyle-markers synergistically increased CRC risk. Of those risk factors, dietary PFA intake and long lifetime exposure to female hormones may play a key role in mediating racial disparity of CRC incidence between AA and HA women.

**Conclusions:** Our results may improve CRC risk prediction performance in those medically/scientifically underrepresented groups and lead to the development of genetically informed interventions for cancer prevention and therapeutic effort, thus contributing to reduced cancer disparities in those minority subpopulations.

**Keywords:** glucose homeostasis, random survival forest, attributable risk, smoking, endogenous estrogen, polyunsaturated fatty acid, colorectal cancer, African and Hispanic/Latino American women

## INTRODUCTION

Although cancer mortality has declined throughout all racial/ethnic groups since 1971 when the National Cancer Act, known as the “War on Cancer”, began, cancer health disparities still exist in the form of higher cancer incidence and mortality among the racial/ethnic minorities (1). In particular, colorectal cancer (CRC) incidence and death rates in African American (AA) women are highest among all racial/ethnic female groups and, compared with white women, 20% and 35%, respectively, were higher during 2012–2016 (2, 3). Also, in the 2 largest minorities, AA and Hispanic/Latino American (HA) women, CRC is the third leading cause of cancer diagnosis and related death (3, 4).

The risk for CRC development increases in older women. For example, approximately 90% of new CRC cases occur in women 50 years old and older (2), and one of the main risk factors is excessive adiposity (5, 6). Specifically, among AA and HA postmenopausal women of at least age 50 years, our preliminary analysis (Table S1) of abdominal adiposity (measured by waist circumference and waist-to-hip ratio) supported the role of obesity in increased risk for CRC, despite insufficient statistical power. For the major biologic mechanism of colorectal tumorigenesis due to obesity, insulin resistance (IR) or glucose intolerance has been thought to play a key mediating role (7, 8). Specifically, increased levels of glucose and insulin, reflecting IR, which interacts with obesity, promoted colorectal epithelial proliferation (9); the elevated insulin levels stimulated the growth of CRC in both cell lines (10) and an animal model (11). IR promotes mitosis by overexpressing insulin receptors and insulin-like growth factor 1 receptors and by dysregulating downstream cellular signaling cascades, resulting in enhancement of cellular anabolic status and increased anti-apoptosis and cell proliferation (12, 13). IR may thus initiate and facilitate CRC cell growth. However, studies focusing on AA and HA women for IR in relation to CRC risk are lacking. One study of DNA methylation in association with CRC among AAs (mainly women) (14) revealed aberrant methylation of CpG islands in the genes that are involved in an insulin network, suggesting the critical role of IR in AA women’s colorectal carcinogenesis. Also, the preliminary results (Table S1) in AA and HA women from our analysis of the fasting glucose and insulin levels (FG and FI) indicated that increased levels of both molecules (particularly glucose) were associated with higher risk for CRC in both groups, but these findings lacked sufficient power to reach significance.

Considering that the systemic development of IR can be influenced by not only environmental (15–17) but also genetic

factors (18, 19), studying genomic markers that explain variations of glucose and insulin concentrations may provide more confirmatory understanding of those concentrations’ role in CRC development. The effort to detect genetic variations of IR has been made in extensive genomic studies, but they mostly focused on whites. AAs and HAs are thus underrepresented in genetic/genome-wide studies of IR. Uncovering IR-specific genetic signatures in these large minorities may advance the understanding of the biology of IR regulation and further, as cancer biomarkers, improve the prediction ability for CRC risk. It can also promote the development of genetically focused, tailored interventions for CRC preventive and therapeutic efforts.

For this reason, we conducted a genomic study of IR and, with validated IR-specific genetic variants, tested for the association with CRC risk specifically focusing on AA and HA postmenopausal women. Since the allele frequencies of modeled genotypes and their effects on IR and CRC are race/ethnicity specific, we conducted our genomic study separately within AA and HA women. We examined more than 780 IR single-nucleotide polymorphisms (SNPs) that have been detected as top genetic signals in the largest and independent genome-wide association (GWA) studies (20–25). With the IR-SNPs validated in our datasets, we tested for the association with CRC development.

Moreover, although obesity is most prevalent in both AA and HA women of all racial/ethnic groups (26), and the diabetes rates within those 2 minority groups are higher than they are in whites (27), CRC incidence is more prevalent in AA women than in HA women (3, 28). Our preliminary analysis also supported this phenomenon [hazard ratio (HR)<sub>HA vs. AA</sub> = 1.85, 95% confidence interval (CI): 1.08 – 3.18] (Figure S1); this suggests the potential role of other lifestyle factors (e.g., diet, smoking, alcohol, female hormones) that are also associated with CRC risk (2, 29–38) in mediating the racial/ethnic differences in CRC risk. Therefore, we incorporated these CRC-associated lifestyle factors with IR genetic markers that we validated for their associations with IR and CRC risk and established risk-prediction models in AA and HA women. By computing the risk prediction for each variable for CRC risk, we detected the most influential genetic markers and lifestyle factors. We next estimated the prediction ability and accuracy of those risk factors, both singly and combined. We further computed to what extent genetic and lifestyle factors, separately and together, influence the development of CRC in each racial/ethnic group [i.e., population attributable risk (PAR)]. Eventually, we estimated an attributable risk (AR) for the common risk factors across the 2 groups to explore potential factors that may play a key role in differentiating the risk for CRC between groups.

## MATERIALS AND METHODS

### Study Subjects

Our study subjects were AA and HA postmenopausal women who had been enrolled in the SNP Health Association Resource (SHARe), which is a prospective cohort of the minorities as a part of Women's Health Initiative Database for Genotypes and Phenotypes (WHI dbGaP) Harmonized and Imputed GWA Studies with the aim of revealing genes/genetic variants in association with quantitative traits with enhanced statistical power in those racial/ethnic minorities. Details of the study design and rationale have been described elsewhere (39–41). In brief, healthy women were recruited at 40 WHI-designated clinical centers across the United States from 1993 through 1998 if they were 50–79 years old, postmenopausal, and expected to stay near the clinical centers for at least 3 years after enrollment. Women were excluded if they had any medical conditions associated with predicted survival of less than 3 years in the judgment of the clinical center physician. They had been further enrolled in the WHI dbGaP study if they had met eligibility for data submission to the dbGaP resource and provided DNA samples. Participants provided written informed consent at enrollment. Among 10,818 women (7,470 AA and 3,348 HA) who reported their race or ethnicity as AA or HA, we applied exclusion criteria as follows: genomic data quality control (QC); a history of diabetes; a diagnosis of any cancer type at enrollment; and less than 1-year follow-up. Ultimately, our study cohort contained 6,678 women (4,692 AA and 1,986 HA). After enrollment, they had been followed through August 2014, with a median follow-up of 15 years at the end point. By their last follow-up, 89 women [73 (1.5%) AA and 16 (0.8%) HA] had developed primary CRC. The institutional review boards of the WHI participating clinical centers and the University of California, Los Angeles approved our study.

### Selection of IR SNPs

We employed data to select IR-specific SNPs from the publicly available genomic resource on glycemic traits, the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC; [www.magicinvestigators.org](http://www.magicinvestigators.org)) (20–23). MAGIC had analyzed FG and FI as continuous variables. We also used 2 other GWA-based data resources for racial/ethnic minorities. One (24) detected SNPs associated with FG in a 500-kb linkage disequilibrium (LD) block, and the other (25) found functional SNPs for glucose intolerance. Among a total of 1,344 FG-SNPs and 313 FI-SNPs identified in these studies, 689 FG and 91 FI SNPs for AA women and 692 FG and 92 FI SNPs for HA women are available in our SHARe dbGaP study, among which 94 FG and 8 FI SNPs for AAs and 168 FG and 1 FI SNPs for HAs were validated with a relevant phenotype.

### Genotyping and Phenotyping

We extracted genotyping data for the study subjects from the WHI dbGaP SHARe database. Details of genotyping information have been reported (39, 41). DNA samples were obtained from the subject blood samples at baseline and genotyped with Affymetrix 6.0 (Affymetrix, Inc., Santa Clara, CA) at the Fred

Hutchinson Cancer Research Center in Seattle, WA. Genomic data were normalized to Genome Reference Consortium Human Build 37, imputed with the 1000 genomes reference panels, and harmonized *via* pairwise concordance among samples across WHI GWA studies. We compared the self-reported ethnicity with genetic principal component (PC). If any discrepancy or admixed participant was found, the subject was labeled as being genetically inconsistent; no one in the SHARe data was identified whose genetic ethnicity was inconsistent. We conducted genomic data QC, filtering out those SNPs with a missing-call rate of  $\geq 2\%$ , a Hardy-Weinberg equilibrium of  $p < 1E-04$ , and  $\check{R}^2 < 0.6$  imputation quality (42). Further, we excluded those individuals with unexpected duplicates, first- and second-degree relatives, and outliers defined by our genetic PC analysis.

Blood samples after fasting were derived from each subject at baseline by trained phlebotomists. Serum levels of glucose and insulin were measured using the hexokinase method on a Hitachi 747 instrument (Boehringer Mannheim Diagnostics, Indianapolis, IN) and using a radioimmunoassay method (Linco Research, Inc., St. Louis, MO), respectively, with average coefficients of variation of 1.28% and 10.93%, respectively.

### Lifestyle Factors and Cancer Outcome

To select CRC-associated lifestyle factors, we performed a literature review (2, 29–38, 43–46) particularly focusing on AAs and HAs. On the basis of our review, we extracted the following lifestyle variables from the SHARe database: age at enrollment; family history of CRC (genetic inheritance); lipid metabolic profiles; anthropometric measures (body mass index [BMI], waist circumference, and waist-to-hip ratio); physical activity; alcohol intake (daily dietary alcohol intake and history of alcohol intake); smoking (number of years as a regular smoker and number of cigarettes smoked daily); nutrition (dietary fiber; daily fruits and vegetables; percent calories from protein; percent calories from saturated and mono- and polyunsaturated fatty acids [SFA, MFA, and PFA, respectively]; dietary calcium; vitamin K; and total sugars); age at menopause; and duration of oral contraceptive (OC) use. Additionally, we included in our data analysis the following variables: demographic and socioeconomic variables (education; marital status; and employment); comorbid conditions (depressive symptoms; cardiovascular disease ever; and hypertension ever); and other reproductive histories (age at menarche; number of pregnancies; duration of breast feeding; oophorectomy and/or hysterectomy; and unopposed/opposed exogenous estrogen use). All the aforementioned variables had been obtained at baseline from subjects *via* self-administered questionnaires, except weight, height, and waist/hip circumferences, which had been measured by trained clinical staff. The WHI coordinating clinical centers monitored all the data collection processes. By using those 35 selected variables, we further conducted preliminary univariate and stepwise/multiple regressions in association with CRC risk and checked multicollinearity between variables.

A diagnosis of primary CRC in the study subjects was confirmed *via* a centralized review of medical records and

pathology and cytology reports by the WHI committee of physicians, who followed the National Cancer Institute's Surveillance, Epidemiology, and End-Results guidelines (47). The time between enrollment and CRC diagnosis, censoring, or study end-point was computed, first in days, and then converted to years.

## Statistical Analysis

We conducted linear and Cox proportional hazards regressions to estimate the relationship of GWA-based IR-SNPs with naturally log-transformed FG (mg/dl)/FI ( $\mu$ IU/ml) and with CRC risk, respectively, after confirming that the assumptions for each were met. Both regression analyses were adjusted for age and 10 genetic PCs that account for racial/ethnic ancestry variations. A 2-tailed  $p < 0.05$  for validation tests of FG/FI and association tests with CRC risk was considered nominally significant. After the Bonferroni correction for multiple comparisons,  $p < 7E-05$  for FG,  $p < 5E-04$  for FI, and  $p < 5E-04$  (in AAs) and  $p < 3E-04$  (in HAs) for CRC risk were considered statistically significant.

With those SNPs validated for their association with relevant phenotype and CRC risk and the selected lifestyle factors, we conducted a Random Survival Forest (RSF) analysis. RSF is a tree-based ensemble machine-learning method that accounts for the nonlinear effects and high-order interactions among variables (48); it has outperformed traditional prediction models, successfully yielding more accurate predictions (49–53). The 2 key predictive values generated from the RSF model are minimal depth (MD); those variables with a small MD are highly predictive, and variable importance (VIMP); those variables with a larger VIMP are more predictive (48, 54). RSF creates a tree from the bootstrapped samples by maximizing survival differences across daughter nodes and, by repeating this process numerous times ( $n = 5,000$  trees in this study), generates a forest of trees. Using the out-of-bag (OOB) data, we first computed the prediction error and next, the OOB concordance index ( $c\text{-index} = 1 - \text{prediction error}$ ), which is conceptually similar to the area under the receiver operating characteristic (ROC) curve (AUC) (55, 56).

We applied a multimodal RSF approach in the AA and HA groups to detect the most influential predictors for CRC risk among the SNPs and lifestyle factors. In a separate RSF analysis within genetic markers and lifestyle variables, we first compared the 2 key predictive values, MD and VIMP, in the plot. Next, we computed the incremental error rate of each variable within the nested sequenced RSF models. Last, we estimated the drop error rate in each variable ranked by MD in the nested models to detect variables that contribute to reducing the prediction error rate. By using the identified topmost influential SNPs and lifestyle factors, both singly and combined in each group of women, we further estimated the OOB  $c\text{-index}$  within the nested RSF model and plotted an ROC curve (57) to quantitatively measure their prediction performance. Further, we estimated the combined effect of the topmost genetic and lifestyle predictors on CRC risk using Cox regression in each racial/ethnic group. After a 2-tailed  $p$  value was corrected for multiple comparisons *via* the Benjamini-Hochberg method, a 5% false discovery rate (FDR)

was considered statistically significant. Eventually, by using the most predictive variables in each group, we computed the PAR percentage (58) to determine the extent to which CRC cases in the group are attributed to genetic and lifestyle factors, singly and in combination. Last, we identified common variables from the most influential variables among the AA and HA women, and by estimating the AR percentage for each variable (59), we explored what variable(s) may contribute to the racial difference in CRC incidence between the groups. Multiple R packages were used (R v4.0.4, pROC survival, survivalROC, randomForestSRC, ggRandomForests, ggplot2, ggthemes, and gamlss).

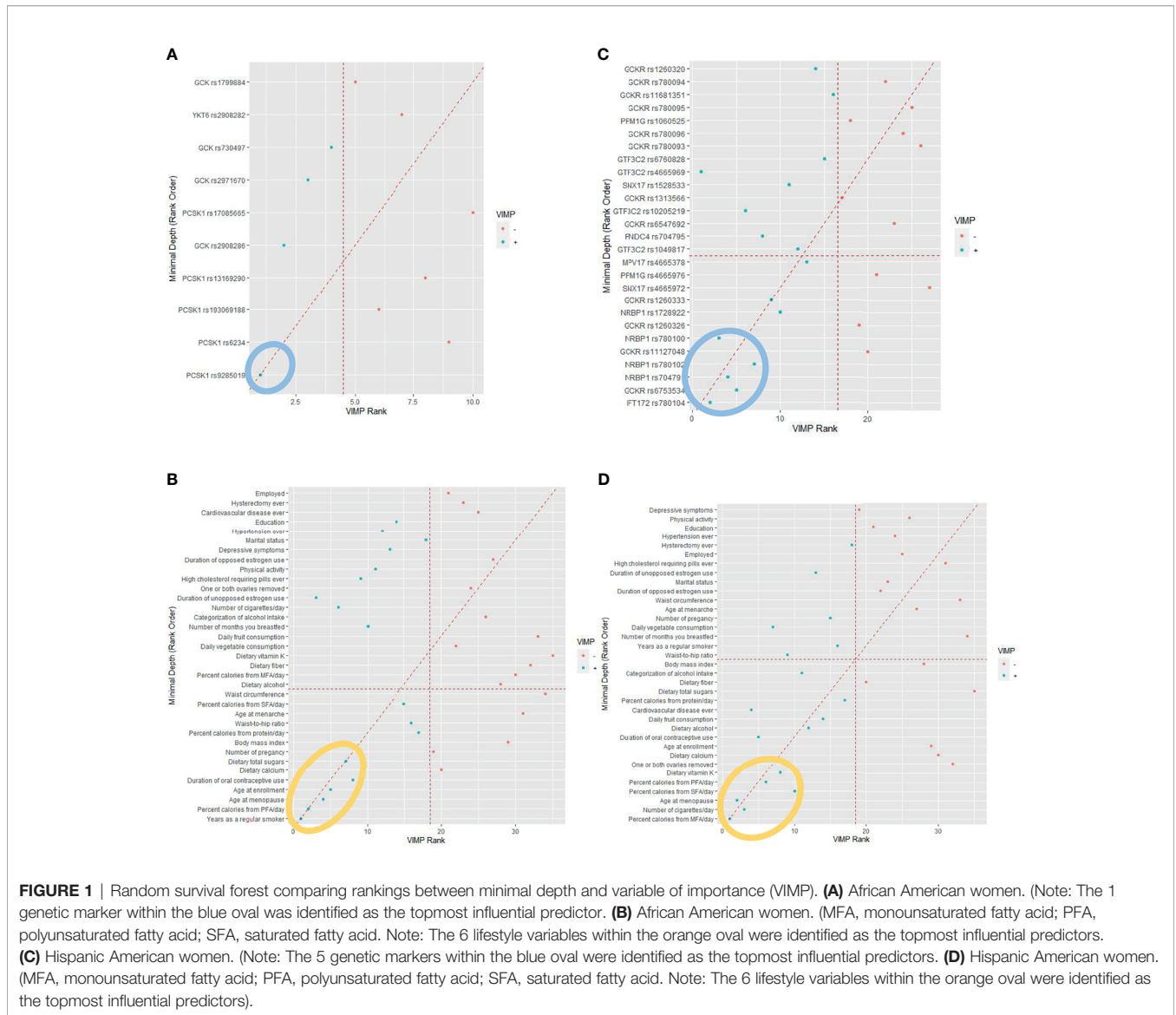
## RESULTS

Between the 94 FG and 8 FI SNPs in AA women (Tables S2A, B) and the 168 FG and 1 FI SNPs (Tables S3A, S2B) in HA women, which were validated with a relevant phenotype nominally and after multiple comparison corrections, 35 FG SNPs overlapped, while none of the FI SNPs were shared by the AA and HA groups. In the analysis of those validated SNPs for their association with the risk of CRC development, 10 SNPs in AA women (Table S2C) and 27 SNPs in HA women (Table S3C) were significant nominally and after multiple comparison correction. Of note, they were all identified among the FG SNPs and were not shared by the 2 groups: the FG SNPs of AAs were from the chromosomes 5 and 7, whereas the FG SNPs of HAs were from chromosome 2. Using those SNPs validated with the phenotype and CRC outcomes in each group of women, we proceeded to the next step, RSF analysis.

### Multimodal RSF Analysis of Validated SNPs and Selected Lifestyle Factors

To detect the topmost influential genetic and lifestyle factors in each racial/ethnic group within the RSF prediction model, we adapted a multimodal approach. In separate RSF models within the SNPs and selected lifestyle factors, we first generated a plot of 2 prediction measures, the MD and VIMP (Figure 1). In agreement with high ranks between the 2 values in AA women, we detected 1 genetic and 6 lifestyle factors as the topmost predictive variables for CRC risk (Figures 1A, B): *PCSK1* rs9285019 and years as a regular smoker, percent calories from PFA/day, dietary total sugar intake, age at enrollment, age at menopause, and duration of OC use. Next, we computed the incremental and drop error rates of each SNP and lifestyle variable arranged by MD in the nested sequenced RSF models (Tables S4A, B), detecting the same set of the topmost 1 genetic and 6 lifestyle variables, which contributes substantially to reducing the prediction error rate. By using these topmost predictive variables, we further estimated a  $c\text{-index}$  and AUC (Table 1) and plotted them (Figure 2A), confirming those top variables' prediction ability. Specifically, in the  $c\text{-index}$  plots for the SNP (Figure 2Aa) and lifestyles (Figure 2Ac), which were ordered by MD rank, those topmost genetic and lifestyle variables were distinctive to improve prediction ability compared with the rest of the variables. The AUC estimations





for those topmost genetic and lifestyle variables each presented results similar to those from the c-index estimation (**Figures 2Ab, 2Ad**). The combination of the gene- and lifestyle-specific AUC yielded 0.647 (95% CI: 0.587 – 0.708) (**Figure 2Ae**), revealing that the topmost lifestyle variables were more substantial contributors to the prediction performance than the top genetic marker was.

We applied the same approach to the group of HA women to find their topmost influential variables. We detected 5 SNPs and 6 lifestyles in agreement with high ranks between MD and VIMP (**Figures 1C, D**) and, by computing the incremental/drop error rate of each genetic and lifestyle variable (**Tables S4C, D**), we identified those same topmost genetic and lifestyle variables. Due to the high LD ( $r^2 > 0.5$ ) within the detected topmost 5 SNPs, we determined 3 SNPs (*IFT172* rs780104, *GCKR* rs6753534, and *NRBP1* rs704791) as the final influential genetic markers and carried them over to the c-index/AUC estimation (**Table 1**

and **Figure 2B**). The topmost lifestyle variables identified in the HA women were similar to those detected in the AA women, but more variables were involved: dietary fat intake (SFA/MFA) and dietary vitamin K intake. The c-index and AUC measures from a separate analysis within these topmost SNPs (**Figures 2Ba, Bb**) and lifestyle factors (**Figures 2Bc, Bd**) also indicated their prediction ability. The AUC from the SNPs and lifestyles together was 0.830 (95% CI 0.721 – 0.939) (**Figure 2Be**), in which those top genetic factors contributed more profoundly to the prediction ability than the top lifestyle factors did; this pattern differs from that observed in AA women.

### The Detected Topmost SNPs and Lifestyle Factors: Combined Effects on CRC Risk

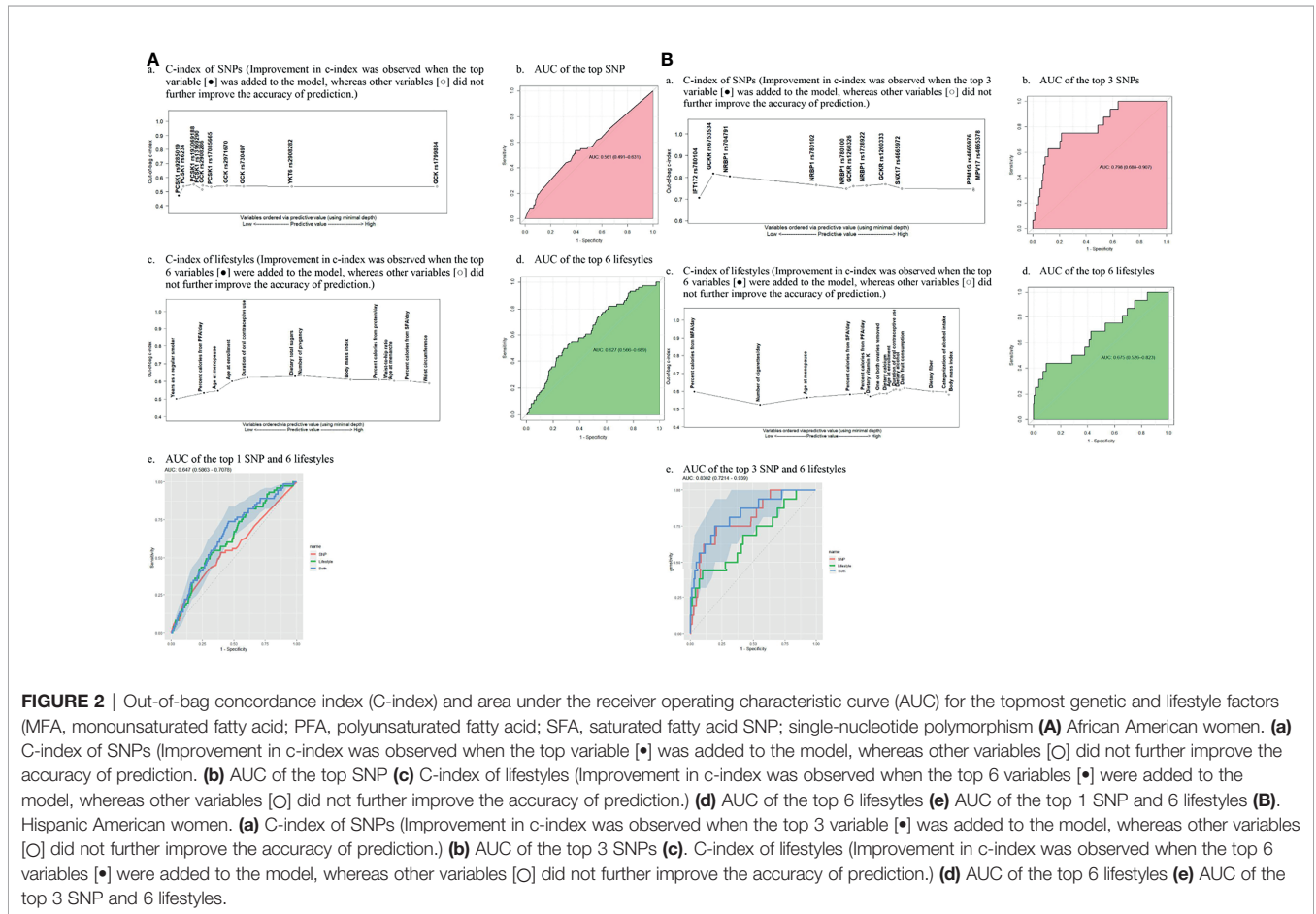
By using the topmost influential IR-SNPs and lifestyle variables in each racial/ethnic group, we implemented the machine-learning process using the RSF model to compute the

**TABLE 1 |** Predictive measures C-index and AUC of the topmost genetic and lifestyle factors in association with colorectal cancer risk.

Type of variable	African American women			Hispanic American women		
	Topmost influential variables*	C-index	AUC (95% CI)	Topmost influential variables*	C-index	AUC (95% CI)
<b>SNP</b>	<i>PCSK1</i> rs9285019	0.4715	0.561 (0.491 – 0.631)	<i>IFT172</i> rs780104 <i>GCKR</i> rs6753534 <i>NRBP1</i> rs704791	0.7064 0.8175 0.8048	0.798 (0.688 – 0.907)
<b>Lifestyle factors</b>	Years as a regular smoker % calories from PFA/day Age at menopause Age at enrollment Duration of OC use Dietary total sugars	0.5023 0.5356 0.5486 0.6014 0.6223 0.6301	0.627 (0.566 – 0.689)	% calories from MFA/day Number of cigarettes/day Age at menopause % calories from SFA/day % calories from PFA/day Dietary vitamin K	0.5979 0.5245 0.5655 0.5836 0.5896 0.5721	0.675 (0.526 – 0.823)
<b>SNP + Lifestyle factors</b>	1 SNP + 6 lifestyle factors		0.647 (0.586 – 0.708)	3 SNPs + 6 lifestyle factors		0.830 (0.721 – 0.939)

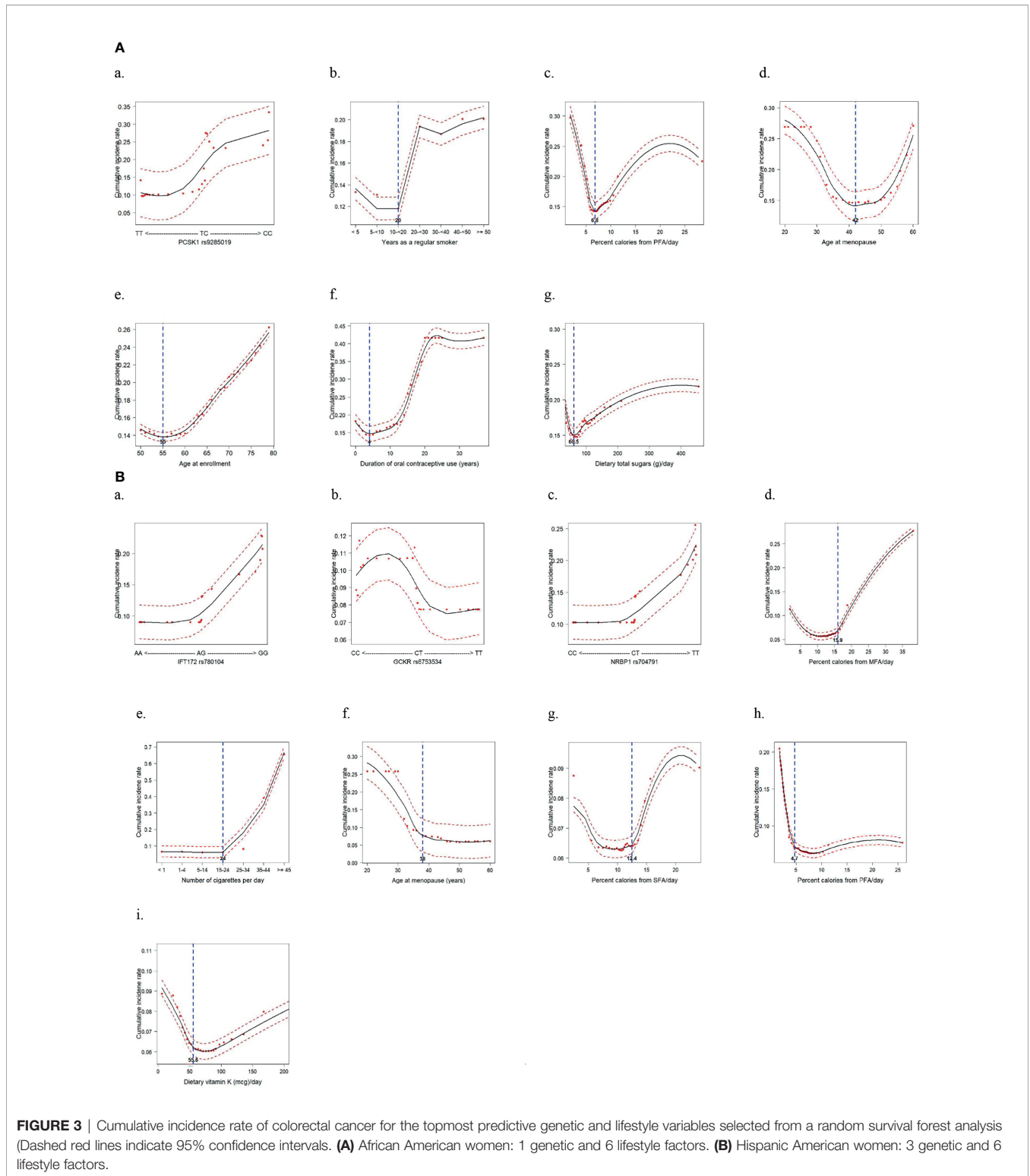
AUC, area under the receiver operating characteristic curve; CI, confidence interval; C-index, concordance index; MFA, monounsaturated fatty acid; OC, oral contraceptive; PFA, polyunsaturated fatty acid; SFA, saturated fatty acid; SNP, single-nucleotide polymorphism.

\*Topmost predictive variables were selected on the basis of random survival forest analysis with a multimodal approach.



cumulative predictive CRC incidence rate by adjusting for confounding variables and a nonlinearity effect of the variable on CRC incidence (Figure 3). In the AA group, the risk genotype and risk lifestyles were defined according to their cutoff values, which were determined by their risk distribution in the plot:

*PCSK1* rs9285019 TC+CC;  $\geq 20$  years as a regular smoker;  $\leq 6.8\%$  of daily calories from PFA; age  $> 42$  years at menopause; age between 56 and 79 years at enrollment; 5–37 years of OC use; and  $> 60.5$  g of total dietary sugar intake. In the HA group, *IFT172* rs780104 GG, *GCKR* rs6753534 CC, and *NRBP1*



rs704791 TT were determined to be the risk genotypes. Also, > 15.9% of daily calories from MFA; ≥ 25 cigarettes smoked daily; age ≤ 38 years at menopause; > 12.4% of daily calories from SFA; ≤ 4.7% of daily calories from PFA; and ≤ 55.6 mg of dietary

vitamin K were defined as the risk lifestyles. It is noteworthy that in both groups, a greater daily intake of calories from PFA was shown to be a protective factor against CRC development. Interestingly, prolonged exposure to female hormones (i.e., late

menopause and/or longer OC use) was revealed to be a risk factor for CRC development among the AA women, but in the HA women it was a protective factor.

Having categorized those topmost SNPs and lifestyle variables accordingly, we first investigated their individual risks for CRC (by adjusting for the others), thus confirming their single effects on CRC risk (**Table S5**). Indeed, the effect magnitude of the individual SNPs was much greater in the HA group than it was in AA group; this corresponded with the finding of the greater influence of those topmost SNPs on the AUC in the HA than in the AA women. Also, whereas most lifestyle variables were not significant after accounting for the others in the AA group, some of the lifestyle variables in the HA group were significant, having a substantial effect on CRC risk.

Next, we tested for the combined effect of the topmost influential SNPs and lifestyle variables, both singly and together, on the risk for CRC. Referring to the analysis of the number of combined lifestyles in relation to CRC risk (**Figure S2A**) in AA women, we combined the AA women with 5 or 6 risk lifestyles and compared their risk with that of the AA women with  $\leq 4$  risk lifestyles. This yielded an approximately 3 times increased risk for CRC in this high risk–lifestyle group (**Table 2**). Further, we combined the risk genotype and lifestyle factors to test for their synergistic effect on increasing risk for CRC. Compared with the women without either of genetic and lifestyle factors, the AA women with both risk factors were associated with a 4-times higher risk for developing CRC, suggesting a gene–lifestyle dose-response relationship in both additive and multiplicative interaction models (HR of G×E = 1.08). In the HA women, stronger effects of SNPs and lifestyle factors, in each combination, were observed (**Table 3**): about 10 times higher risk for CRC among those with 2 or 3 risk alleles than among those with none or 1 risk allele; and about 7 times greater CRC risk among those with 3 risk lifestyles than among those with  $\leq 2$  risk lifestyles. The

maximum number of lifestyle combinations was 3, and they were categorized on the basis of CRC risk distribution by the number of combined lifestyles (**Figure S2B**). Consistent with our findings of the AA women, the HA women who had both risk genotypes and risk lifestyles had greater and much stronger ( $> 58$  times) risk for CRC than did those who did not have either of them (**Table 3**). This also suggests that the most-predictive genetic and lifestyle factors in combination synergistically increased the predictability of CRC risk in both additive and multiplicative interaction models (HR of G×E = 1.38).

## PAR Percentage for the Combined Topmost Variables in Each Group and AR Percentage for the Variables Common to Both Groups

In the estimation of PAR percentage from the topmost genetic and lifestyle variables in AA women, 23% of their CRC cases were attributed to one top SNP, and 33% were attributed to lifestyle factors in combination. Further, 45% of the CRC cases in AA women were attributed to those genetic and lifestyle factors combined, implicating that almost half of the cases could have been prevented if they would not have had such risk factors (**Table 2**). In HA women, 67% of the CRC cases was attributed to genetic factors, and 26% was attributed to risk lifestyles. When the top genetic and lifestyle factors were combined, about 70% of the CRC cases could have been prevented if they had not possessed such risk factors (**Table 3**).

In addition, we detected 3 common lifestyle factors among the topmost influential markers shared by the AA and HA women: smoking, age at menopause, and daily calorie intake from PFA (**Table 4**). The AR percentages from smoking between the groups were similar, but those from age at menopause and dietary PFA intake were 2 times and 4 times higher, respectively, in the HA than they were in the AA women. The

**TABLE 2 |** African American women: combined effect of risk genotypes and risk lifestyles on colorectal cancer risk and population-attributable risk percentage.

Number of risks	n	HR (95% CI)	p	PAR (%) <sup>†</sup>
<b>Risk genotypes<sup>‡</sup></b>				
0	2,756	reference		22.9
1	1,936	<b>1.64 (1.03 – 2.59)</b>	<b>0.0356</b>	
<b>Risk lifestyles<sup>§</sup></b>				
0	3,097	reference		33.6
1	1,595	<b>2.61 (1.65 – 4.15)</b>	<b>4.66E-05</b>	
<b>Risk genotypes plus lifestyle factors<sup>§</sup></b>				
0	1,859	reference		44.9
Risk genotypes only	1,238	1.51 (0.75 – 3.02)	0.2450	
Risk lifestyles only	897	<b>2.46 (1.25 – 4.83)</b>	<b>0.0088*</b>	
Both risks of genotypes and lifestyles	698	<b>4.02 (2.12 – 7.60)</b>	<b>1.95E-05*</b>	
$\rho$ trend			<b>1.00E-04</b>	

CI, confidence interval; HR, hazard ratio; PAR, population attributable risk. Numbers in bold face are statistically significant.

<sup>†</sup>PAR(%) reflects, in total African American women, a risk of colorectal cancer attributable to the risk genotypes and the risk lifestyles, both singly and in combination.

<sup>‡</sup>The number of risk genotype (PCSK1 rs9285019 TC+CC) was defined as follows: 0 (none) vs. 1 (1 risk allele).

<sup>§</sup>The number of lifestyles ( $\geq 20$  years as a regular smoker,  $\leq 6.8\%$  of calories from polyunsaturated fatty acid/day,  $> 42$  years old at menopause, 56–79 years old at enrollment, 5–37 years of oral contraceptive use, and  $> 60.5$  g of dietary total sugars) was determined on the basis of analysis for the combined lifestyle factors (**Figure S2A**) and defined as follows: 0 (null/1/2/3/4 risk lifestyles) vs. 1 (5/6 risk lifestyles).

<sup>§</sup>The combined number of risk genotypes and risk lifestyles was based on risk genotype defined as 0 (none) and 1 (1 risk allele), and risk lifestyles defined as 0 (null/1/2/3/4 risk lifestyles) and 1 (5/6 risk lifestyles). The ultimate number of risk genotypes combined with risk lifestyles was defined as 0 (no risk genotypes and risk lifestyles); and risk genotypes (only risk genotypes) and risk lifestyles (only risk lifestyles), separately and together.

\*p values with false discovery rate  $< 0.05$  are shown after multiple comparison corrections via the Benjamini-Hochberg method.



**TABLE 3** | Hispanic American women: combined effect of risk genotypes and risk lifestyles on colorectal cancer risk and population-attributable risk percentage.

Number of risks	n	HR (95% CI)	p	PAR (%) <sup>†</sup>
<b>Risk genotypes<sup>‡</sup></b>				
0	1,495	reference		66.8
1	491	<b>9.57 (3.08 – 29.67)</b>	<b>9.20E-05</b>	
<b>Risk lifestyles<sup>¶</sup></b>				
0	1,850	Reference		26.2
1	136	<b>6.63 (2.30 – 19.11)</b>	<b>0.0005</b>	
<b>Risk genotypes plus lifestyle factors<sup>§</sup></b>				
0	1,394	Reference		
Risk genotypes only	456	<b>8.55 (2.27 – 32.24)</b>	<b>1.53E-03*</b>	73.3
Risk lifestyles only	101	4.97 (0.52 – 47.76)	0.1653	
Both risks of genotypes and lifestyles	35	<b>58.76 (13.15 – 262.68)</b>	<b>9.73E-08*</b>	
$\rho$ trend			<b>2.00E-06</b>	

CI, confidence interval; HR, hazard ratio; PAR, population attributable risk. Numbers in bold face are statistically significant.

<sup>†</sup>PAR(%) reflects, in total Hispanic African women, a risk of colorectal cancer attributable to the risk genotypes and the risk lifestyles, both singly and in combination.

<sup>‡</sup>The number of risk genotypes (IFT172 rs780104 GG; GCKR rs6753534 CC; and NRBP1 rs704791 TT) was defined as follows: 0 (none/1 risk allele) vs. 1 (2/3 risk alleles).

<sup>¶</sup>The maximum combined number of lifestyles (> 15.9% of calories from monounsaturated fatty acid [FA]/day,  $\geq 25$  cigarettes/day,  $\leq 38$  years old at menopause, > 12.4% of calories from saturated FA/day,  $\leq 4.7\%$  of calories from polyunsaturated FA/day, and  $\leq 55.6$  mg of dietary vitamin K) was 3. The number of lifestyles was determined on the basis of analysis for the combined lifestyle factors (Figure S2B) and defined as follows: 0 (null/1/2 risk lifestyles) vs. 1 (3 risk lifestyles).

<sup>§</sup>The combined number of risk genotypes and risk lifestyles was based on risk genotypes defined as 0 (none/1 risk allele) and 1 (2/3 risk alleles), and risk lifestyles defined as 0 (null/1/2 risk lifestyles) and 1 (3 risk lifestyles). The ultimate number of risk genotypes combined with risk lifestyles was defined as 0 (no risk genotypes and risk lifestyles); and risk genotypes (only risk genotypes) and risk lifestyles (only risk lifestyles), separately and together.

\*p values with false discovery rate < 0.05 were shown after multiple comparison corrections via the Benjamini-Hochberg method.

**TABLE 4** | Colorectal cancer attributable risk for the lifestyle factors detected as the topmost predictive variables in both African American and Hispanic American women.

Overlapped variables:the topmost predictors	African American Women AR (%)	Hispanic American women AR (%)
Smoking <sup>†</sup>	61.7	87.4
Age at menopause	28.2	57.1
percent calories from PFA/day	12.7	48.9

AR, attributable risk; PFA, polyunsaturated fatty acid.

<sup>†</sup>The modeled variable for smoking factor is years as a regular smoker in African American women and the number of cigarettes smoked daily in Hispanic American women.

HA women's long lifetime exposure to female hormones tended to be protective, and the threshold of daily PFA intake to prevent CRC risk was less than the AA women's (5% vs. 7%, respectively). Altogether, we postulate that these 2 lifestyle factors play an important role in mediating the difference in CRC risk between AA and HA women.

## DISCUSSION

Despite some improvement in healthcare disparities between different racial/ethnic categories in cancer medicine, disparities in cancer genomic science still exist for AA and HA women, the 2 largest minorities of the U.S. population, which are underrepresented in collection, aggregation, and analysis of genomic data for studies of cancer risk factors. Here we focused on AA and HA postmenopausal women to examine genetic markers of IR, one of the main biologic mechanisms of colorectal carcinogenesis, by using an extensive set of GWA-based IR SNPs. In addition to these genetic factors, by incorporating CRC-associated lifestyle variables to establish the CRC risk prediction model for each racial/ethnic group, we detected the topmost influential genetic and lifestyle factors. The combined topmost genetic- and lifestyle-specific markers revealed a synergistic effect on increasing the CRC risk by

explaining a considerable portion of their cancer risk. Thus, constructing CRC risk profiles with those topmost markers substantially improved the risk-prediction performance. We believe that these results could be used in the development of genetically focused interventions for cancer prevention and therapeutic effort, and allow progress toward reducing cancer disparity in those minorities.

Most of the topmost FG-SNPs we detected are found in the intronic and intergenic regions of genes that play well-established roles in modulating glucose metabolism, implicating that these genetic variations may influence glucose homeostasis. In AA women, the genetic variant in the *PCSK1* gene was associated with FG concentration as well as increased risk for CRC. The *PCSK1* gene encodes prohormone convertase 1/3, which mediates the cleavage of proinsulin in the process of insulin biosynthesis. Thus, that gene mutation leads to the loss-of-function defect in insulin production, eventually resulting in impaired glucose tolerance (60–63). Further, the mutation of this gene is associated with carcinogenesis and enhanced cancer growth, particularly in the liver metastasis of primary CRC cells (64), suggesting the involvement of the convertases in the selective process of liver metastasis. To the best of our knowledge, ours is the first report of the *PCSK1* gene variation's association with primary CRC risk, particularly in AA women.

Of the topmost FG-SNPs detected in HA women, the genetic variant of *GCKR* was associated with a higher FG concentration and increased CRC risk. The *GCKR* regulates the activity of glucokinase in liver and pancreatic islet cells (65). For example, when circulating glucose level is low, *GCKR* forms an inactive complex with glucokinase, inhibiting glycolysis (66). Thus, a high degree of inhibition of this enzyme by *GCKR* can result in high FG levels. The genetic variation of *GCKR* in association with FG concentrations was previously reported in AAs (24) but not in HAs. Also, the *GCKR* variation has been associated with the risk of pancreatic cancer (67) and the prognosis of metastatic gastric cancer (68), but no published study so far has examined its association with CRC risk. Therefore, our findings of FG and CRC risk in HA women are meaningful and warrant replication in further studies with independent datasets. In addition, *NRBP1*, which encodes multidomain putative adapter proteins (69), has an anti-tumor role against CRC tumorigenesis and progression, as an *in vivo/in vitro* study (70) showed that the higher expression of *NRBP1* inhibited CRC cell proliferation and anti-apoptosis and correlated with better prognosis. *NRBP1* regulates the apoptotic pathway by inhibiting *Jab1*-mediated *JNK* signaling, which is essential in gene translation and regulation of cellular apoptosis (70–72); it may thus play a key role in suppressing CRC tumorigenesis. Supported by these earlier findings, our study reported that the variation of the *NRBP1* gene increased the risk of CRC, specifically in HA women. Last, the genetic variants of *IFT172* that encodes a subunit of the intraflagellar transport subcomplex *IFT-B*, which is necessary for ciliary assembly and maintenance, have been associated with ulcerative colitis and Crohn's disease (73), but their associations with CRC risk, as detected in our study, have not been previously reported, warranting future replication studies.

Among the 3 topmost influential factors shared by the AA and HA groups, the effect of smoking on CRC risk was strongest in both groups. As revealed in a recent Mendelian randomization study (31), prolonged lifetime exposure to cigarette smoking is positively associated with CRC risk. The carcinogens emitted by tobacco smoke into the digestive system and bloodstream promote tumorigenesis in colorectal mucosa (74). In particular, AA individuals tend to have higher total equivalents of nicotine per number of cigarettes smoked daily than individuals of other racial/ethnic groups, and their CRC screening rate is lower in active smokers than in never smokers (75); thus, screening in the high-risk group (active/longer-term regular smokers) is strongly recommended.

Both groups in our study had greater risk for CRC when they had lower daily intake of PFAs. Previous studies (29, 76) support our finding, by reporting that the decreased proportions of red blood cell PFAs and less intake of PFAs were associated with increased CRC incidence. PFAs have been shown to suppress pro-inflammatory cytokine production (77) and reduce triglycerides and low-density lipoprotein particles (78), which are key mediators in carcinogenesis. In our HA women, the CRC risk attributable to low PFA intake was more substantial than it

was in our AA women. However, the HA women had a lower threshold of daily PFA intake than AA women in preventing CRC development. Altogether, the effect of less strict requirement of PFA intake in HA women may override their more sensitive influence of low PFA intake on CRC risk and thus, contribute to the lower CRC incidence in HA than in AA women.

Further, older age at menopause is an important risk factor for CRC development in postmenopausal women (79–81), suggesting that longer lifetime exposure to endogenous estrogen may increase the CRC risk. However, in our analysis of HA women, their longer-term exposure to female hormones tended to be protective against CRC risk, even after adjusting for a history of oophorectomy; this suggests a follow-up functional mechanism study in this racial/ethnic subpopulation. Similar to that of PFA intake, this protective role of prolonged lifetime exposure to female hormones in HA women may outweigh the greater effect of short-term hormone exposure on CRC risk than AA women had, explaining in part their lower CRC incidence compared with that of AA women.

Our data on smoking were self-reported, so our results may have been subject to misclassification bias. However, a previous study found high reliability of self-reported assessment of active smoking (82). Also, our RSF analysis may overfit the model with multiple tasks, warranting the conduct of replication studies with independent datasets. We examined AA and HA postmenopausal women, so our findings may not be generalizable to other racial/ethnic populations.

Overall, our study indicates that GWA-level IR SNPs combined with the lifestyle factors of smoking, lifetime exposure to endogenous female hormones, and dietary fat intake synergistically increased the risk for CRC, and the prediction ability and accuracy of these factors was notable. Of those risk factors, dietary intake of PFAs and lifelong exposure to female hormones may play a key role in mediating the racial disparity of CRC risk between AA and HA women. Our findings may improve CRC risk–prediction performance in these medically and scientifically underrepresented subpopulations, and by emphasizing the promotion of genetically informed preventive interventions (e.g., smoking cessation, higher PFA intake) and encouraging CRC screening of individuals who are at high risk owing to particular risk genotypes and behavioral patterns, our results may contribute to reduced cancer disparity in those minorities.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: The data that support the findings of this study are available in accordance with policies developed by the NHLBI and WHI in order to protect sensitive participant information and approved by the Fred Hutchinson Cancer Research Center, which currently serves as the IRB of record for the WHI. Data requests may be made by emailing [helpdesk@WHI.org](mailto:helpdesk@WHI.org).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional review boards of each participating clinical center of the WHI and the University of California, Los Angeles. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

SJ, ES, MP, HY, and JP designed the study. SJ performed the genomic data QC. SJ performed the statistical analysis and SJ, ES, MP, HY, and JP interpreted the data. JP and ES supervised the genomic data QC and analysis and participated in the study coordination. JP oversaw the project. SJ secured funding for this project. All participated in writing and editing the paper. All authors contributed to the article and approved the submitted version.

## FUNDING

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

## ACKNOWLEDGMENTS

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).

**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.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2021.760243/full#supplementary-material>

## REFERENCES

- Ramirez AG, Thompson IM. How Will the 'Cancer Moonshot' Impact Health Disparities? *Cancer Causes Control* (2017) 28(9):907–12. doi: 10.1007/s10552-017-0927-6
- American Cancer Society. *Colorectal Cancer Facts & Figures 2020-2021*. Atlanta: American Cancer Society, Inc (2020). Available at: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2020-2022.pdf>.
- American Cancer Society. *Cancer Fact and Figures for African Americans 2019-2021*. Atlanta: American Cancer Society, Inc (2021). Available at: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-facts-and-figures-for-african-americans/cancer-facts-and-figures-for-african-americans-2019-2021.pdf>.
- American Cancer Society. *Cancer Fact and Figures for Hispanics/Latinos 2018-2020*. Atlanta: American Cancer Society, Inc (2018). Available at: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-facts-and-figures-for-hispanics-and-latinos/cancer-facts-and-figures-for-hispanics-and-latinos-2018-2020.pdf>.
- Ma Y, Yang Y, Wang F, Zhang P, Shi C, Zou Y, et al. Obesity and Risk of Colorectal Cancer: A Systematic Review of Prospective Studies. *PLoS One* (2013) 8(1):e53916. doi: 10.1371/journal.pone.0053916
- Shokrani B, Brim H, Hydari T, Afsari A, Lee E, Nouriae M, et al. Analysis of Beta-Catenin Association With Obesity in African Americans With Premalignant and Malignant Colorectal Lesions. *BMC Gastroenterol* (2020) 20(1):274. doi: 10.1186/s12876-020-01412-x
- Abdelsatir AA, Husain NE, Hassan AT, Elmadhoun WM, Almobarak AO, Ahmed MH. Potential Benefit of Metformin as Treatment for Colon Cancer: The Evidence So Far. *Asian Pac J Cancer Prev* (2015) 16(18):8053–8. doi: 10.7314/apjcp.2015.16.18.8053
- Ho GY, Wang T, Gunter MJ, Strickler HD, Cushman M, Kaplan RC, et al. Adipokines Linking Obesity With Colorectal Cancer Risk in Postmenopausal Women. *Cancer Res* (2012) 72(12):3029–37. doi: 10.1158/0008-5472.CAN-11-2771
- Tran TT, Naigamwalla D, Oprescu AI, Lam L, McKeown-Eyssen G, Bruce WR, et al. Hyperinsulinemia, But Not Other Factors Associated With Insulin Resistance, Acutely Enhances Colorectal Epithelial Proliferation *In Vivo*. *Endocrinol* (2006) 147(4):1830–7. doi: 10.1210/en.2005-1012
- Bjork J, Nilsson J, Hultcrantz R, Johansson C. Growth-Regulatory Effects of Sensory Neuropeptides, Epidermal Growth Factor, Insulin, and Somatostatin on the non-Transformed Intestinal Epithelial Cell Line IEC-6 and the Colon Cancer Cell Line HT 29. *Scand J gastroenterol* (1993) 28(10):879–84. doi: 10.3109/00365529309103129
- Tran TT, Medline A, Bruce WR. Insulin Promotion of Colon Tumors in Rats. *Cancer epidemiol Biomarkers Prev Publ Am Assoc Cancer Res cosponsored by Am Soc Prev Oncol* (1996) 5(12):1013–5.
- Sandhu MS, Dunger DB, Giovannucci EL. Insulin, Insulin-Like Growth Factor-I (IGF-I), IGF Binding Proteins, Their Biologic Interactions, and Colorectal Cancer. *J Natl Cancer Inst* (2002) 94(13):972–80. doi: 10.1093/jnci/94.13.972

13. Mulholland HG, Murray LJ, Cardwell CR, Cantwell MM. Glycemic Index, Glycemic Load, and Risk of Digestive Tract Neoplasms: A Systematic Review and Meta-Analysis. *Am J Clin Nutr* (2009) 89(2):568–76. doi: 10.3945/ajcn.2008.26823
14. Ashktorab H, Darempouran M, Goel A, Varma S, Leavitt R, Sun X, et al. DNA Methylation Profiling Identifies Novel Methylated Genes in African American Patients With Colorectal Neoplasia. *Epigenet* (2014) 9(4):503–12. doi: 10.4161/epi.27644
15. 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(2):745–52. doi: 10.3892/ijo.2012.1480
16. Liu J, Carnero-Montoro E, van Dongen J, Lent S, Nedeljkovic I, Ligthart S, et al. An Integrative Cross-Omics Analysis of DNA Methylation Sites of Glucose and Insulin Homeostasis. *Nat Commun* (2019) 10(1):2581. doi: 10.1038/s41467-019-10487-4
17. Franks PW, Mesa JL, Harding AH, Wareham NJ. Gene-Lifestyle Interaction on Risk of Type 2 Diabetes. *Nutr Metab Cardiovasc Dis NMCD*. (2007) 17(2):104–24. doi: 10.1016/j.numecd.2006.04.001
18. Arner P, Sahlqvist AS, Sinha I, Xu H, Yao X, Waterworth D, et al. The Epigenetic Signature of Systemic Insulin Resistance in Obese Women. *Diabetologia* (2016) 59(11):2393–405. doi: 10.1007/s00125-016-4074-5
19. Jung SY, Mancuso N, Yu H, Papp J, Sobel E, Zhang ZF. Genome-Wide Meta-Analysis of Gene-Environmental Interaction for Insulin Resistance Phenotypes and Breast Cancer Risk in Postmenopausal Women. *Cancer Prev Res (Phila)* (2019) 12(1):31–42. doi: 10.1158/1940-6207.CAPR-18-0180
20. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New Genetic Loci Implicated in Fasting Glucose Homeostasis and Their Impact on Type 2 Diabetes Risk. *Nat Genet* (2010) 42(2):105–16. doi: 10.1038/ng.520
21. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-Scale Association Analyses Identify New Loci Influencing Glycemic Traits and Provide Insight Into the Underlying Biological Pathways. *Nat Genet* (2012) 44(9):991–1005. doi: 10.1038/ng.2385
22. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A Genome-Wide Approach Accounting for Body Mass Index Identifies Genetic Variants Influencing Fasting Glycemic Traits and Insulin Resistance. *Nat Genet* (2012) 44(6):659–69. doi: 10.1038/ng.2274
23. Lagou V, Magi R, Hottenga JJ, Grallert H, Perry JRB, Bouatia-Naji N, et al. Sex-Dimorphic Genetic Effects and Novel Loci for Fasting Glucose and Insulin Variability. *Nat Commun* (2021) 12(1):24. doi: 10.1038/s41467-020-19366-9
24. Ramos E, Chen G, Shriner D, Doumatey A, Gerry NP, Herbert A, et al. Replication of Genome-Wide Association Studies (GWAS) Loci for Fasting Plasma Glucose in African-Americans. *Diabetologia* (2011) 54(4):783–8. doi: 10.1007/s00125-010-2002-7
25. Mondal AK, Sharma NK, Elbein SC, Das SK. Allelic Expression Imbalance Screening of Genes in Chromosome 1q21-24 Region to Identify Functional Variants for Type 2 Diabetes Susceptibility. *Physiol Genomics* (2013) 45(13):509–20. doi: 10.1152/physiolgenomics.00048.2013
26. Goding Sauer A, Siegel RL, Jemal A, Fedewa SA. Current Prevalence of Major Cancer Risk Factors and Screening Test Use in the United States: Disparities by Education and Race/Ethnicity. *Cancer Epidemiol Biomarkers Prev Publ Am Assoc Cancer Res Cosponsored by Am Soc Prev Oncol* (2019) 28(4):629–42. doi: 10.1158/1055-9965.EPI-18-1169
27. Bolen JC, Rhodes L, Powell-Griner EE, Bland SD, Holtzman D. State-Specific Prevalence of Selected Health Behaviors, by Race and Ethnicity—Behavioral Risk Factor Surveillance System, 1997. *MMWR CDC Surveill Summ*. (2000) 49(2):1–60.
28. DeLellis K, Rinaldi S, Kaaks RJ, Kolonel LN, Henderson B, Le Marchand L. Dietary and Lifestyle Correlates of Plasma Insulin-Like Growth Factor-1 (IGF-1) and IGF Binding Protein-3 (IGFBP-3): The Multiethnic Cohort. *Cancer Epidemiol Biomarkers Prev Publ Am Assoc Cancer Res Cosponsored by Am Soc Prev Oncol* (2004) 13(9):1444–51.
29. Linseisen J, Grundmann N, Zoller D, Kuhn T, Jansen E, Chajes V, et al. Red Blood Cell Fatty Acids and Risk of Colorectal Cancer in The European Prospective Investigation Into Cancer and Nutrition (EPIC). *Cancer Epidemiol Biomarkers Prev Publ Am Assoc Cancer Res cosponsored by Am Soc Prev Oncol* (2021) 30(5):874–85. doi: 10.1158/1055-9965.EPI-20-1426
30. Sarkissyan M, Wu Y, Chen Z, Mishra DK, Sarkissyan S, Giannikopoulos I, et al. Vitamin D Receptor FokI Gene Polymorphisms May be Associated With Colorectal Cancer Among African American and Hispanic Participants. *Cancer* (2014) 120(9):1387–93. doi: 10.1002/cncr.28565
31. Dimou N, Yarmolinsky J, Bouras E, Tsilidis KK, Martin RM, Lewis SJ, et al. Causal Effects of Lifetime Smoking on Breast and Colorectal Cancer Risk: Mendelian Randomization Study. *Cancer Epidemiol Biomarkers Prev Publ Am Assoc Cancer Res cosponsored by Am Soc Prev Oncol* (2021) 30(5):953–64. doi: 10.1158/1055-9965.EPI-20-1218
32. Bohorquez M, Sahasrabudhe R, Criollo A, Sanabria-Salas MC, Velez A, Castro JM, et al. Clinical Manifestations of Colorectal Cancer Patients From a Large Multicenter Study in Colombia. *Med (Baltimore)* (2016) 95(40):e4883. doi: 10.1097/MD.0000000000004883
33. Reilly MP, Rader DJ. The Metabolic Syndrome: More Than the Sum of Its Parts? *Circulation* (2003) 108(13):1546–51. doi: 10.1161/01.CIR.0000088846.10655.E0
34. Murphy N, Strickler HD, Stanczyk FZ, Xue X, Wassertheil-Smoller S, Rohan TE, et al. A Prospective Evaluation of Endogenous Sex Hormone Levels and Colorectal Cancer Risk in Postmenopausal Women. *J Natl Cancer Inst* (2015) 107(10):1–10. doi: 10.1093/jnci/djv210
35. Lavasani S, Chlebowski RT, Prentice RL, Kato I, Wactawski-Wende J, Johnson KC, et al. Estrogen and Colorectal Cancer Incidence and Mortality. *Cancer* (2015) 121(18):3261–71. doi: 10.1002/cncr.29464
36. Manson JE, Chlebowski RT, Stefanick ML, Aragaki AK, Rossouw JE, Prentice RL, et al. Menopausal Hormone Therapy and Health Outcomes During the Intervention and Extended Poststopping Phases of the Women's Health Initiative Randomized Trials. *JAMA* (2013) 310(13):1353–68. doi: 10.1001/jama.2013.278040
37. Slattery ML, Potter JD, Curtin K, Edwards S, Ma KN, Anderson K, et al. Estrogens Reduce and Withdrawal of Estrogens Increase Risk of Microsatellite Instability-Positive Colon Cancer. *Cancer Res* (2001) 61(1):126–30.
38. Issa JP. Colon Cancer: It's CIN or CIMP. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2008) 14(19):5939–40. doi: 10.1158/1078-0432.CCR-08-1596
39. NCBI. *WHI Harmonized and Imputed GWAS Data. A Sub-Study of Women's Health Initiative* (2019). Available at: [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000746.v3.p3](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000746.v3.p3).
40. Design of the Women's Health Initiative Clinical Trial and Observational Study. The Women's Health Initiative Study Group. *Control Clin Trials* (1998) 19(1):61–109. doi: 10.1016/s0197-2456(97)00078-0
41. NCBI. *Women's Health Initiative - SNP Health Association Resource. A Sub-Study of Women's Health Initiative* (2021). Available at: [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000386.v8.p3](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000386.v8.p3).
42. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association Analyses of More Than 140,000 Men Identify 63 New Prostate Cancer Susceptibility Loci. *Nat Genet* (2018) 50(7):928–36. doi: 10.1038/s41588-018-0142-8
43. Lathroum L, Ramos-Mercado F, Hernandez-Marrero J, Villafana M, Cruz-Correa M. Ethnic and Sex Disparities in Colorectal Neoplasia Among Hispanic Patients Undergoing Screening Colonoscopy. *Clin Gastroenterol Hepatol* (2012) 10(9):997–1001. doi: 10.1016/j.cgh.2012.04.015
44. Centers for Disease C. Prevention. Monthly Estimates of Leisure-Time Physical Inactivity—United States, 1994. *MMWR Morb Mortal Wkly Rep* (1997) 46(18):393–7.
45. He J, Stram DO, Kolonel LN, Henderson BE, Le Marchand L, Haiman CA. The Association of Diabetes With Colorectal Cancer Risk: The Multiethnic Cohort. *Br J Cancer* (2010) 103(1):120–6. doi: 10.1038/sj.bjc.6605721
46. Colbert LH, Hartman TJ, Malila N, Limburg PJ, Pietinen P, Virtamo J, et al. Physical Activity in Relation to Cancer of the Colon and Rectum in a Cohort of Male Smokers. *Cancer Epidemiol Biomarkers Prev Publ Am Assoc Cancer Res cosponsored by Am Soc Prev Oncol* (2001) 10(3):265–8.
47. National Cancer Institute. *SEER Program: Comparative Staging Guide For Cancer* (1993). Available at: [https://seer.cancer.gov/archive/manuals/historic/comp\\_stage1.1.pdf](https://seer.cancer.gov/archive/manuals/historic/comp_stage1.1.pdf).
48. Mogensen UB, Ishwaran H, Gerds TA. Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. *J Stat Software* (2012) 50(11):1–23. doi: 10.18637/jss.v050.i11
49. Chung RH, Chen YE. A Two-Stage Random Forest-Based Pathway Analysis Method. *PLoS One* (2012) 7(5):e36662. doi: 10.1371/journal.pone.0036662



50. Montazeri M, Beigzadeh A. Machine Learning Models in Breast Cancer Survival Prediction. *Technol Health Care Off J Eur Soc Eng Med* (2016) 24 (1):31–42. doi: 10.3233/THC-151071
51. 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(16):2028–36. doi: 10.1093/bioinformatics/btl344
52. Chang JS, Yeh RF, Wiencke JK, Wiemels JL, Smirnov I, Pico AR, et al. Pathway Analysis of Single-Nucleotide Polymorphisms Potentially Associated With Glioblastoma Multifforme Susceptibility Using Random Forests. *Cancer epidemiol Biomarkers Prev Publ Am Assoc Cancer Res cosponsored by Am Soc Prev Oncol* (2008) 17(6):1368–73. doi: 10.1158/1055-9965.EPI-07-2830
53. Tong X, Feng Y, Li JJ. Neyman-Pearson Classification Algorithms and NP Receiver Operating Characteristics. *Sci adv* (2018) 4(2):ea01659. doi: 10.1126/sciadv.aao1659
54. 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(2):250–9. doi: 10.1161/CIRCULATIONAHA.111.058719
55. Ishwaran H, Kogalur UB. *Random Survival Forests for R* (2007). Available at: <https://pdfs.semanticscholar.org/951a/840f176076fb6786df43320e8b27094dcfa.pdf>.
56. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random Survival Forests. *Ann Appl Stat* (2008) 2(3):841–60. doi: 10.1214/08-AOAS169
57. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves. *BMC Bioinf* (2011) 12:77. doi: 10.1186/1471-2105-12-77
58. W Kirch ed. Population Attributable Risk (PAR) Population Attributable Risk (PAR). In: *Encyclopedia of Public Health*. Dordrecht: Springer Netherlands. p. 1117–8.
59. W Kirch ed. Attributable Risk Proportion Attributable Risk Proportion. In: *Encyclopedia of Public Health*. Dordrecht: Springer Netherlands. p. 54–4.
60. Kaufmann JE, Irminger JC, Mungall J, Halban PA. Proinsulin Conversion in GH3 Cells After Coexpression of Human Proinsulin With the Endoproteases PC2 and/or PC3. *Diabetes* (1997) 46(6):978–82. doi: 10.2337/diab.46.6.978
61. Bailyes EM, Shennan KI, Seal AJ, Smeekens SP, Steiner DF, Hutton JC, et al. A Member of the Eukaryotic Subtilisin Family (PC3) has the Enzymic Properties of the Type 1 Proinsulin-Converting Endopeptidase. *Biochem J* (1992) 285(Pt 2):391–4. doi: 10.1042/bj2850391
62. Ramos-Molina B, Martin MG, Lindberg I. PCSK1 Variants and Human Obesity. *Prog Mol Biol Transl Sci* (2016) 140:47–74. doi: 10.1016/bs.pmbts.2015.12.001
63. Stijnen P, Ramos-Molina B, O'Rahilly S, Creemers JW. PCSK1 Mutations and Human Endocrinopathies: From Obesity to Gastrointestinal Disorders. *Endocr Rev* (2016) 37(4):347–71. doi: 10.1210/er.2015-1117
64. Tzimas GN, Chevet E, Jenna S, Nguyen DT, Khatib AM, Marcus V, et al. Abnormal Expression and Processing of the Proprotein Convertases PC1 and PC2 in Human Colorectal Liver Metastases. *BMC Cancer* (2005) 5:149. doi: 10.1186/1471-2407-5-149
65. Sagen JV, Odili S, Bjorkhaug L, Zelent D, Buettger C, Kwagh J, et al. From Clinicogenetic Studies of Maturity-Onset Diabetes of the Young to Unraveling Complex Mechanisms of Glucokinase Regulation. *Diabetes* (2006) 55(6):1713–22. doi: 10.2337/db05-1513
66. Beer NL, Tribble ND, McCulloch LJ, Roos C, Johnson PR, Orho-Melander M, et al. The P446L Variant in GCKR Associated With Fasting Plasma Glucose and Triglyceride Levels Exerts its Effect Through Increased Glucokinase Activity in Liver. *Hum Mol Genet* (2009) 18(21):4081–8. doi: 10.1093/hmg/ddp357
67. Prizment AE, Gross M, Rasmussen-Torvik L, Peacock JM, Anderson KE. Genes Related to Diabetes may be Associated With Pancreatic Cancer in a Population-Based Case-Control Study in Minnesota. *Pancreas* (2012) 41(1):50–3. doi: 10.1097/MPA.0b013e3182247625
68. Liu X, Chen Z, Zhao X, Huang M, Wang C, Peng W, et al. Effects of IGF2BP2, KCNQ1 and GCKR Polymorphisms on Clinical Outcome in Metastatic Gastric Cancer Treated With EOF Regimen. *Pharmacogenomics* (2015) 16(9):959–70. doi: 10.2217/pgs.15.49
69. Hooper JD, Baker E, Ogbourne SM, Sutherland GR, Antalis TM. Cloning of the cDNA and Localization of the Gene Encoding Human NRBP, a Ubiquitously Expressed, Multidomain Putative Adapter Protein. *Genomics* (2000) 66(1):113–8. doi: 10.1006/geno.2000.6167
70. Liao Y, Yang Z, Huang J, Chen H, Xiang J, Li S, et al. Nuclear Receptor Binding Protein 1 Correlates With Better Prognosis and Induces Caspase-Dependent Intrinsic Apoptosis Through the JNK Signalling Pathway in Colorectal Cancer. *Cell Death Dis* (2018) 9(4):436. doi: 10.1038/s41419-018-0402-7
71. Wang H, Sun X, Luo Y, Lin Z, Wu J. Adapter Protein NRBP Associates With Jab1 and Negatively Regulates AP-1 Activity. *FEBS Lett* (2006) 580(25):6015–21. doi: 10.1016/j.febslet.2006.10.002
72. Yarza R, Vela S, Solas M, Ramirez MJ. C-Jun N-Terminal Kinase (JNK) Signaling as a Therapeutic Target for Alzheimer's Disease. *Front Pharmacol* (2015) 6:321. doi: 10.3389/fphar.2015.00321
73. *Gene Card: Human Gene Database: IFT172 Gene (Protein Coding)*. Available at: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=IFT172>.
74. Yamasaki E, Ames BN. Concentration of Mutagens From Urine by Absorption With the Nonpolar Resin XAD-2: Cigarette Smokers Have Mutagenic Urine. *Proc Natl Acad Sci United States America* (1977) 74(8):3555–9. doi: 10.1073/pnas.74.8.3555
75. Oluyemi AO, Welch AR, Yoo LJ, Lehman EB, McGarrity TJ, Chuang CH. Colorectal Cancer Screening in High-Risk Groups Is Increasing, Although Current Smokers Fall Behind. *Cancer* (2014) 120(14):2106–13. doi: 10.1002/cncr.28707
76. Williams CD, Satia JA, Adair LS, Stevens J, Galanko J, Keku TO, et al. Associations of Red Meat, Fat, and Protein Intake With Distal Colorectal Cancer Risk. *Nutr Cancer* (2010) 62(6):701–9. doi: 10.1080/01635581003605938
77. Pischon T, Hankinson SE, Hotamisligil GS, Rifai N, Willett WC, Rimm EB. Habitual Dietary Intake of N-3 and N-6 Fatty Acids in Relation to Inflammatory Markers Among US Men and Women. *Circulation* (2003) 108(2):155–60. doi: 10.1161/01.CIR.0000079224.46084.C2
78. Griffin MD, Sanders TA, Davies IG, Morgan LM, Millward DJ, Lewis F, et al. Effects of Altering the Ratio of Dietary N-6 to N-3 Fatty Acids on Insulin Sensitivity, Lipoprotein Size, and Postprandial Lipemia in Men and Postmenopausal Women Aged 45–70 Y: The OPTILIP Study. *Am J Clin Nutr* (2006) 84(6):1290–8. doi: 10.1093/ajcn/84.6.1290
79. Zervoudakis A, Strickler HD, Park Y, Xue X, Hollenbeck A, Schatzkin A, et al. Reproductive History and Risk of Colorectal Cancer in Postmenopausal Women. *J Natl Cancer Inst* (2011) 103(10):826–34. doi: 10.1093/jnci/djr101
80. Talamini R, Franceschi S, Dal Maso L, Negri E, Conti E, Filiberti R, et al. The Influence of Reproductive and Hormonal Factors on the Risk of Colon and Rectal Cancer in Women. *Eur J Cancer* (1998) 34(7):1070–6. doi: 10.1016/S0959-8049(98)00019-7
81. Yoo KY, Tajima K, Inoue M, Takezaki T, Hirose K, Hamajima N, et al. Reproductive Factors Related to the Risk of Colorectal Cancer by Subsite: A Case-Control Analysis. *Br J Cancer* (1999) 79(11-12):1901–6. doi: 10.1038/sj.bjc.6690302
82. Soulakova JN, Hartman AM, Liu B, Willis GB, Augustine S. Reliability of Adult Self-Reported Smoking History: Data From the Tobacco Use Supplement to the Current Population Survey 2002-2003 Cohort. *Nicotine Tob Res* (2012) 14(8):952–60. doi: 10.1093/ntr/ntr313

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a shared affiliation, though no other collaboration, with several of the authors SJ, ES, MP, JP.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Jung, Sobel, Pellegrini, Yu and Papp. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.