Am J Cancer Res 2020;10(9):2955-2976
www.ajcr.us /ISSN:2156-6976/ajcr0119567

Original Article

Pro-inflammatory cytokine polymorphisms in ONECUT2 and HNF4A and primary colorectal carcinoma: a post genome-wide gene-lifestyle interaction study

Su Yon Jung[1], Jeanette C Papp[2], Eric M Sobel[2,3], Matteo Pellegrini[4], Herbert Yu[5], Zuo-Feng Zhang[6,7]

[1]Translational Sciences Section, Jonsson Comprehensive Cancer Center, School of Nursing, University of California, Los Angeles, CA 90095, USA; [2]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA; [3]Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA; [4]Department of Molecular, Cell and Developmental Biology, Life Sciences Division, University of California, Los Angeles, CA 90095, USA; [5]Cancer Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI 96813, USA; [6]Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, CA 90095, USA; [7]Center for Human Nutrition, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

Received August 4, 2020; Accepted August 12, 2020; Epub September 1, 2020; Published September 15, 2020

Abstract: Immune-related molecular and genetic pathways that are connected to colorectal cancer (CRC) and lifestyles in postmenopausal women are incompletely characterized. In this study, we examined the role of pro-inflammatory biomarkers such as C-reactive protein (CRP) and interleukin-6 (IL-6) in those pathways. Through selection of the best predictive single-nucleotide polymorphisms (SNPs) and lifestyles, our goal was to improve the prediction accuracy and ability for CRC risk. Using large cohort data of postmenopausal women from the Women's Health Initiative Database for Genotypes and Phenotypes Study, we previously conducted a genome-wide association (GWA) for a CRP and IL-6 gene-behavioral interaction study. For the present study, we added GWA-SNPs from outside GWA studies, resulting in a total of 152 SNPs. Together with 41 selected lifestyles, we performed a 2-stage multimodal random survival forest analysis with generalized multifactor dimensionality reduction approach to construct CRC risk profiles. Overall and in obesity strata (by body mass index, waist circumference, waist-to-hip ratio, exercise, and dietary fat intake), we identified the best predictive genetic markers in inflammatory cytokines and lifestyles. Across the strata, 2 SNPs (*ONECUT2* rs4092465 and *HNF4A* rs1800961) and 1 lifestyle factor (relatively short-term past use of oral contraceptives) were the most common and strongest predictive markers for CRC risk. The risk profile that combined those variables exhibited synergistically increased risk for CRC; this pattern appeared more strongly in obese and inactive subgroups. Our results may contribute to improved predictability for CRC and suggest genetically targeted lifestyle interventions for women carrying the inflammatory-risk genotypes, reducing CRC risk.

Keywords: Random survival forest, generalized multifactor dimensionality reduction, inflammatory cytokines, C-reactive protein, interleukin-6, oral contraceptive, endogenous estrogen, obesity, colorectal cancer, postmenopausal women

Introduction

Chronic inflammation is a critical factor involved in the pathogenesis of obesity-attributable cancers such as colorectal cancer (CRC) from tumor initiation to progression [1, 2]. In particular, CRC in postmenopausal women ages 50 years and older accounts for most (approximately 90%) of newly diagnosed CRC patients and related deaths [3], contributing to the third ranking of CRC in cancer incidence and mortality among women of the United States and Westernized countries [4, 5]. CRC is indeed an inflammatory-associated disease, being seen as high risk among individuals with inflammatory bowel disease [6] or high levels of the inflammatory cytokines such as C-reactive protein (CRP) and interleukin-6 (IL-6) [7-11]. These pro-inflammatory cytokines reflect different molecular pathways in acute and chronic

immune responses but may be interconnected in carcinogenesis as shown in *in vitro* and *in vivo* studies [2, 12-16]. In detail, CRP, a major acute-phase reactant and biomarker of chronic low-grade inflammation partially induced by IL-6, elevates the mRNA expression of genes (e.g., *LOX-1*) important to CRC development in CRC cell lines [12]. IL-6, upregulated by macrophages and adipose tissue, itself exhibits high mRNA expression and immunoreactivity in the CRC tissue, epithelium, and stroma [13]. Further, both inflammatory markers alter the gut microbiome, gradually forming a microenvironment that is essential to colorectal carcinogenesis [14-16]. In addition, both biomarkers exhibit higher plasma levels or mRNA expression in adipose tissue of CRC patients [17, 18] and in the colon of obese individuals, accompanied by precancerous changes in the transcriptome [19], indicating that obesity-modified inflammatory pathways are associated with colorectal tumorigenesis.

Thus, the genetic variants involved in those biomarkers' functional and structural regulation are potentially implicated in the causal pathway of inflammatory-associated CRC development that interacts with obesity factors. Earlier genomic epidemiology studies for associations between CRP/IL-6-related genetic variants and CRC risk are limited and showed inconsistent results [20-24], with only a slight effect on CRC risk [25]. The gene-phenotype pathways of CRP and IL-6 may not link to each other alone, but may also be connected to lifestyle pathways, thus being modulated by obesity (overall and visceral) [26-34], obesity-related lifestyles including lipid metabolism [34, 35], a high-fat diet, exercise level, smoking, and alcohol consumption [26, 36-43]. Our previous genome-wide association (GWA) study [44] of CRP and IL-6 addressed this pleiotropic effect of those biomarkers on the gene-phenotype relationship and revealed greater relationships between single-nucleotide polymorphisms (SNPs) and phenotypes among overall and viscerally obese, physically inactive, and high dietary-fat subgroups, suggesting the role of obesity in regulating the inflammatory gene-phenotype pathway. Further, the genomic pathways between CRP/IL-6 and CRC can be modified by obesity factors, suggesting that CRP/IL-6 (genotypes and phenotypes) in conjunction with obesity influence the risk of CRC (Figure S1, yellow lines). For this reason, studying how obesity factors modify the effects of genes and phenotypes, contributing to increased CRC susceptibility, is important to promote genetically targeted interventions in primary cancer prevention efforts. However, no studies thus far have examined these key pro-inflammatory cytokines in relation to CRC at the genome-wide level by incorporating a wide range of obesity-related lifestyle factors and different fat distributions of obesity status.

We hoped to address these gaps by focusing on postmenopausal women, a population vulnerable to inflammation [45], obesity, and CRC. By using a large cohort of postmenopausal women from the Women's Health Initiative Database for Genotypes and Phenotypes (WHI dbGaP) Study, we previously performed a GWA gene-environment (G×E) interaction study with meta-analysis of subGWAs for CRP and IL-6 by evaluating interactions with obesity factors. We identified a total of 88 top GWA SNPs overall and in obesity strata [44]. In the present study, we have extended the scope of modeled SNPs by adding another 68 SNPs associated with CRP and IL-6 from earlier GWA studies that focused on European ancestry with independent replications [28, 29, 46, 47]. We examined the relationships of these top GWA SNPs with primary CRC risk overall and in the obesity strata in which the SNPs were initially identified for the association with CRP/IL-6 in our previous GWA G×E study. This allowed us to elucidate a hypothetical empirical pathway in which a substantial proportion of the GWA SNPs in CRP and IL-6 affects CRC risk through interactions with specific lifestyle factors, and thus may contribute to understanding of the genomic immune-related etiologic pathways connected to CRC and lifestyles.

In this study, we hoped to improve the CRC prediction accuracy and ability by better characterizing the genetic architecture of the inflammatory biomarkers with incorporation of modifiable and non-modifiable risk factors. We examined the top GWA SNPs together with 41 selected lifestyles by performing a 2-stage multimodal random survival forest (RSF) analysis and generated their predictive value and accuracy for CRC. RSF, a nonparametric tree-based ensemble learning method, is one of the prediction models that has outperformed the tradi-

tional models in terms of clinical accuracy [48-52]. In particular, the RSF accounts for nonlinearity and high-order interactions among variables [53, 54] and thus may provide a more accurate risk estimation. Further, we applied a generalized multifactor dimensionality reduction (GMDR) approach and characterized high-order gene-gene interactions, selecting the best genetic prediction model [55-58]. With the strongest influential SNPs and lifestyle factors selected by the RSF and GMDR, we finally constructed prediction models for CRC and calculated the combined and joint effects of genotypes and lifestyles on CRC development. We ultimately tested the hypothesis that the most-predictive genetic and lifestyle factors in combination synergistically increase the predictability of CRC risk.

Material and methods

Study population

The present study included healthy postmenopausal women from the WHI Harmonized and Imputed GWA Studies (GWASs) that were coordinated to contribute to a joint effort of imputation and harmonization for GWASs within the 2 WHI study arms, Clinical Trials and Observational Studies. Details of the study designs and rationale are described elsewhere [59, 60], but briefly, healthy women were enrolled in the WHI study from 1993 through 1998 at 40 clinical centers across the U.S. if they met the following eligibility criteria: 50-79 years old, postmenopausal, expected to stay near the clinical centers for at least 3 years after enrollment, and able to provide written informed consent. Participants were further eligible for the WHI dbGaP study if they had met eligibility requirements for submission to the dbGaP and provided DNA samples. Under the dbGaP accession (phs000200.v12.p3), the Harmonization and Imputation GWASs consist of 6 subGWASs (Tables S1A and S1B). In our previous GWA G×E study, we initially included 16,088 women who reported their race or ethnicity as non-Hispanic white (Figure S2) and, by applying the exclusion criteria (i.e., diabetes history; genetic data duplications; first- and second-degree relatives; and genetic quality control [QC] based on principal components), left 10,798 women. In the present study, additional exclusion was made for those with a follow-up period of less than 1 year and/or any cancer diagnosis present at screening, leaving a total of 10,142 women (94% of the eligible women in the GWAS). These women had been followed through August 29, 2014, with a mean of 16 years of follow-up; 737 (7%) of them had developed primary CRC. The Institutional Review Boards of each WHI-participating clinical center and the University of California, Los Angeles, approved this study.

Data collection and CRC outcome

The WHI coordinating centers had collected participant information via self-administered questionnaire and periodically performed data quality assurance. For the purpose of our analysis, we initially selected 41 variables assessed at screening on the basis of their association with inflammation and CRC through a literature review [1, 2, 17-19, 26, 34-43, 61-64], followed by preliminary analyses such as univariate and stepwise multiple regressions and a multicollinearity test. Those examined variables are as follows: demographic (age, education, and marital status) and socioeconomic factors (family income and employment); family history of diabetes and CRC; medical (hypertension, high cholesterol, cardiovascular disease, and depressive symptoms) and reproductive histories (hysterectomy, one or both ovaries removed, ages at menarche and menopause, pregnancy, breast feeding, oral contraceptive [OC] use, and exogenous estrogen [E] only and E plus progestin [E+P] use); lifestyles (physical activity and cigarette smoking); and daily diet (dietary energy, alcohol, total sugar, fiber, fruit, and vegetable consumption; percentage of calories from protein, carbohydrates, saturated fatty acids [SFA], monounsaturated FA [MFA], and polyunsaturated FA [PFA]). The anthropometric variables height, weight, and waist and hip circumferences had been measured by trained staff and were further included in our analysis.

The CRC outcomes were determined via a centralized review of medical charts and pathology/cytology reports by a committee of physicians. Cancer sites were recorded according to the National Cancer Institute's Surveillance, Epidemiology, and End-Results guidelines [65]. The time from enrollment to primary CRC development, censoring, or study end point was measured and calculated in years.

Genotyping and laboratory methods

Genotyped data were extracted from the WHI dbGaP GWASs database, normalized to the reference panel GRCh37, and imputed using 1000 Genomes reference panels [60]. Details of the data-cleaning process have been previously discussed [44, 60, 66]. Briefly, the SNPs' harmonization was checked via pairwise concordance in all samples across the GWASs. The initial data QC step filtered SNPs with a missing call rate of > 2% and a Hardy-Weinberg equilibrium of P < 1E-04. In the second QC step, SNPs with $\hat{R}^2 \geq 0.6$ imputation quality were included [67], but individuals with a KING kinship estimate > 0.088 were excluded [68].

Blood samples from participants who had fasted for at least 8 hours had been collected at baseline by trained phlebotomists. Serum levels of CRP were analyzed via a high-sensitivity immunoturbidimetric assay (Kamiya Biomedical Company) and of IL-6 by the Quantitative Sandwich Enzyme Immunoassay technique (Quantikine HS Immunoassay Kit; R&D Systems, Inc., Minneapolis, MN), with median inter-assay coefficients of variation of 2.3% and 12.4%, respectively.

Statistical analysis

Participants' characteristics at baseline and allele frequencies by CRC were examined via unpaired 2-sample *t* tests (for continuous variables) and chi-squared tests (for categorical variables). If continuous variables were skewed or had outliers, the Wilcoxon rank-sum test was conducted. Our earlier GWAS evaluated obesity factors as an effect modifier via a formal interaction test and stratifications defined by body mass index (BMI; 30 kg/m^2 cutoff), waist circumference (WST; 88 cm), waist-to-hip ratio (WHR; 0.85), metabolic equivalents (METs; 10·hours/week), and percentage of calories from SFA (9%). The testing results from each sub-GWAS were combined via a meta-analysis assuming a fixed effects model. In the present study, we excluded 1 SNP from the 88 SNPs in our previous study due to high missing proportion and examined their association with CRC in a particular lifestyle setting in which the SNPs were identified. Of another 68 SNPs obtained from other GWASs, 3 SNPs were also excluded due to high missing values and were analyzed both overall and in subgroups.

We performed the RSF analysis on the quality adjusted data. The RSF uses bootstrapping to generate samples using about 60% of the original dataset and grows a tree from each sample via a splitting rule to maximize differences in survival rate across daughter nodes. The tree-building process is repeated numerous times (n=5,000 in this study) and aggregated in a forest of trees for prediction [48, 69]. An ensemble cumulative hazard estimate was computed from each tree and averaged over all trees for each individual, and that estimate was used to calculate a predicted cumulative CRC incidence rate. By using this ensemble estimate and creating the out-of-bag (OOB) data from the remaining 40% of the original data, the OOB concordance index (C-index) was calculated, which is a measure of prediction performance similar to the area under the receiver operating characteristic (AUROC) curve [69, 70]. Each variable's ranking was determined according to its prediction ability for CRC via 2 prediction parameters: 1) minimal depth (MD), in which variables that have a small MD split the tree close to the root, thus being considered highly predictive and 2) variable importance (VIMP), on the basis of permutation strategy using the OOB C-index, in which variables that have greater VIMPs are more predictive [53].

A 2-stage RSF was performed (**Figure 1**). In the first stage, we conducted separate RSF analyses on SNPs (overall and in obesity strata) and lifestyles. With only those SNPs and lifestyles that had a significantly low MD and high VIMP, we performed the second RSF with a multimodal approach: overall and in obesity strata 1) comparing MD and VIMP values in the plot, 2) calculating the OOB C-index from the nested RSF model, and 3) computing the incremental error rate of each variable in the nested sequence RSF models and from the top variable, calculating a dropping error rate. This approach allowed us to exclude from the outset the SNPs and lifestyles that were not significantly associated with CRC, resulting in increased statistical power and an adjusted type I error rate [48]. Additionally, we applied a GMDR approach, which is described elsewhere [55-57]. Briefly, the GMDR reduces high-dimensional multifactor prediction to a single dimension by the ratio of high vs. low risk, and it detects the best gene-gene interaction model. It generates key prediction parameters, such

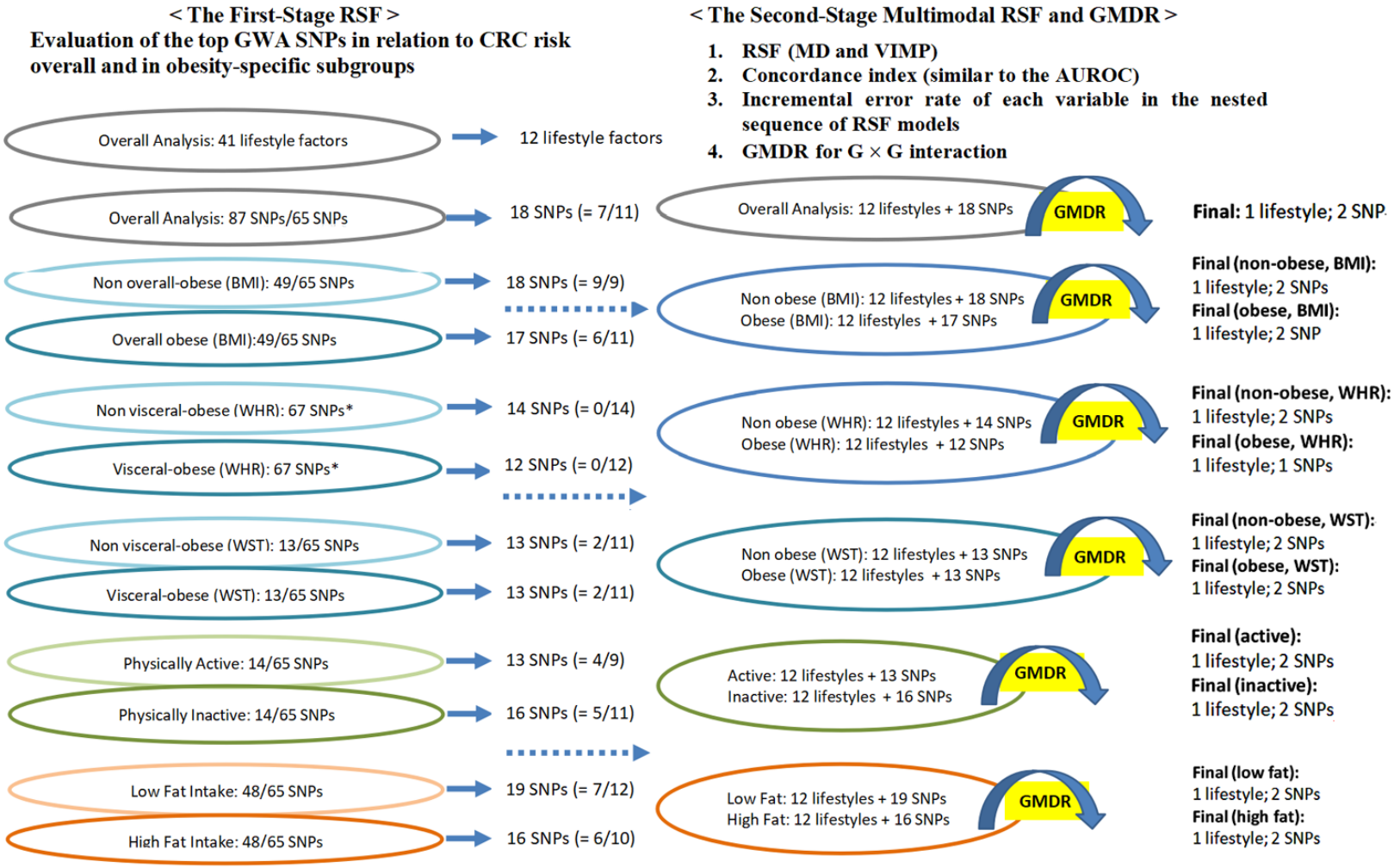


Figure 1. Two-stage random survival forest (RSF) and generalized multifactor dimensionality reduction (GMDR). (BMI, body mass index; CRC, colorectal cancer; GWA, genome-wide association; MD, minimal depth; SNP, single-nucleotide polymorphism; WHR, waist-to-hip ratio; WST, waist circumference; VIMP, variable of importance. * WHR subgroups combined 2 our GWA and 65 outside GWA SNPs).

as testing balance accuracy (TBA), cross-validation consistency (CVC), and sign *p* value. The model with the highest TBA, CVC 10/10, and P < 0.05 based on 1,000-times permutation testing was considered the best model.

We further performed multiple Cox proportional hazards regressions with an assumption test via a Schoenfeld residual plot and r evaluation to obtain hazard ratios (HRs) and 95% confidence intervals (CIs) for the single and combined effects of SNPs and lifestyle factors on CRC, by adjusting for covariates (**Table 1**). A 2-tailed *p* value < 0.05 was considered statistically significant with multiple-comparison corrections by the Benjamini-Hochberg method [71]. R v.3.6.3. (survival, survivalROC, randomForestSRC, ggRandomForests, gamlss, ggsurvplot, and forestplot packages) and GMDR v.1.0. were used.

Results

Participants' characteristics at screening according to CRC development (**Table 1**) and the allele frequencies of 152 GWA CRP/IL-6 SNPs for each WHI subGWAS (Tables S1A and S1B) are displayed. Participants who developed CRC were likely to be younger, highly educated, more depressed, and taller. They also tended to have shorter breastfeeding periods and shorter durations of past OC use but higher frequencies and longer durations of E-only and E+P use.

Two-stage multimodal RSF and GMDR

We analyzed the 152 GWA SNPs (one set of 87 from our GWAS and another set of 65 from outside GWASs) and 41 selected lifestyle factors by implementing the 2-stage RSF and GMDR (**Figure 1**) to identify the most predictive genetic and lifestyle markers with the highest predictability and lowest prediction error for CRC risk. In the first stage of RSF (Figure S3), we calculated the 2 predicted values, MD and VIMP, and plotted to compare them; they use different prediction algorithms, so we expected the variables' ranking to be somewhat different. Separately, for each set of GWA SNPs and the 41 lifestyles, we identified the best predictive genetic and lifestyle factors on the basis of the agreement with high ranks in both MD and VIMP (**Figures 1** and S3) as follows: 12 of 41 lifestyles; 18 of 152 SNPs overall (in detail, 7 from our GWAs and 11 from outside GWASs); 18 and 17 of 114 SNPs (BMI < 30 and ≥ 30, respectively); 14 and 12 of 67 SNPs (WHR ≤ 0.85 and > 0.85); 13 and 13 of 78 SNPs (WST ≤ 88 and > 88); 13 and 16 of 79 SNPs (METs ≥ 10 and < 10); and 19 and 16 of 113 SNPs (SFA < 9 and ≥ 9).

The 12 lifestyles and selected SNPs together, overall and in the obesity strata, were carried over to the second stage of multimodal RSF to generate risk profiles with the most predictive variables that account for both genetic and lifestyle factors. Particularly, in the overall group, we first calculated the 2 measures MD and VIMP (**Table 2**) and plotted them (**Figure 2A**); the dashed red line reflects the agreement of the 2 measures. Both measures in agreement with high ranks revealed that 2 SNPs (*HNF4A* rs1800961 and *ONECUT2* rs4092465) and 1 lifestyle (past OC use) were the best predictive variables for CRC risk. Next, the C-index (i.e., the AUROC) was computed from the nested RSF model (**Table 2**) and plotted (**Figure 2B**), where variables were ranked by MD. From that, we identified the same set of top variables (2 SNPs and 1 lifestyle). Those top variables substantially improved the C-index prediction accuracy (AUROC=0.95), whereas others did not, suggesting the complementary prediction ability of the C-index. Further, we estimated a dropping error rate for each variable in the nested sequence of the RSF models (**Table 2**), identifying once again the same top 3 variables as the strongest contributors that dropped the error rate, thus substantially improving the prediction accuracy. Finally, we applied the GMDR approach (Table S2) and determined the best gene-by-gene interaction model; the 2-factor model including the top 2 SNPs was the most predictive, with the highest TBA of 0.7019 and CVC of 10/10 (P < 0.001).

We continued to apply those 2-stage multimodal RSF (Tables S3A, S3B, S3C, S3D, S3E, S3F, S3G, S3H, S3I, S3J and Figures S4, S5, S6, S7, S8) and GMDR (Table S2) methods to each of the obesity strata (BMI, WHR, WST, MET, and SFA), and throughout the strata, determined the same top 3 variables detected from the overall analysis as the best predictive markers, except in the viscerally obese subgroup (WHR > 0.85), in which only 1 SNP (*ONECUT2* rs4092465) was identified as the best marker.

Table 1. Characteristics of participants, stratified by CRC

Characteristic	Participants without CRC (n=9,405)		Participants with CRC (n=737)	
	n	(%)	n	(%)
Age in years, mean (SD)	67	(6.66)	66	(6.57)*
Education				
≤ High school	3,409	(36.2)	220	(29.9)*
> High school	5,996	(63.8)	517	(70.1)
Marital status				
Currently not married	3,649	(38.8)	288	(39.1)
Currently married	5,756	(61.2)	449	(60.9)
Family income				
< $35,000	4,253	(45.2)	315	(42.7)
≥ $35,000	5,152	(54.8)	422	(57.3)
Employment				
Currently employed (full- or part-time)	6,900	(73.4)	523	(71.0)
Currently not employed	2,505	(26.6)	214	(29.0)
Cardiovascular disease ever				
No	7,995	(85.0)	610	(82.8)
Yes	1,410	(15.0)	127	(17.2)
Hypertension ever				
No	6,506	(69.2)	487	(66.1)
Yes	2,899	(30.8)	250	(33.9)
Family history of CRC				
No	7,932	(84.3)	610	(82.8)
Yes	1,473	(15.7)	127	(17.2)
Depressive symptom†				
< 0.06	8,724	(92.8)	666	(90.4)*
≥ 0.06	681	(7.2)	71	(9.6)
METs·hour·week^{-1}, mean (SD)¶	10.95	(12.84)	11.49	(12.73)
METs·hour·week^{-1}¶				
≥ 10.0	3,881	(41.3)	322	(43.7)
< 10.0	5,524	(58.7)	415	(56.3)
Cigarettes smoked per day				
Never smoked	5,018	(53.4)	380	(51.6)
< 15	2,386	(25.4)	185	(25.1)
≥ 15	2,001	(21.3)	172	(23.3)
Years of regular smoking				
Never smoked	5,018	(53.4)	380	(51.6)
< 5	510	(5.4)	43	(5.8)
5-9	524	(5.6)	37	(5.0)
10 +	3,353	(35.7)	277	(37.6)
Dietary alcohol per day in g, mean (SD)	6.12	(11.45)	6.35	(11.92)
% calories from carbohydrates, mean (SD)	49.01	(8.69)	49.37	(8.91)
Dietary total sugars in g, mean (SD)	98.88	(43.73)	100.02	(42.93)
% calories from SFA, median (range)	11.34	(2.22-32.39)	11.44	(2.60-26.77)
% calories from SFA€				
< 9.0 %	2,121	(22.6)	169	(22.9)

≥ 9.0 %	7,284	(77.4)	568	(77.1)
% calories from MFA, mean (SD)	12.71	(3.26)	12.56	(3.22)
% calories from PFA, mean (SD)	6.83	(2.09)	6.80	(2.04)
Height in cm, mean (SD)	161.9	(6.00)	162.5	(6.07)*
Weight in kg, mean (SD)	72.99	(14.79)	72.98	(14.10)
BMI in kg/m^2, mean (SD)	27.80	(5.36)	27.61	(5.19)
BMI¥				
< 30.0	6,613	(70.3)	538	(73.0)
≥ 30.0	2,792	(29.7)	199	(27.0)
Waist circumference in cm, mean (SD)	86.72	(12.86)	87.3	(12.30)
Waist circumference¥				
≤ 88	5,574	(59.3)	424	(57.5)
> 88	3,831	(40.7)	313	(42.5)
Hip circumference in cm, mean (SD)	106.5	(11.16)	106.6	(10.87)
Waist-to-hip ratio, mean (SD)	0.813	(0.073)	0.8184	(0.074)
Waist-to-hip ratio¥				
≤ 0.85	6,702	(71.3)	520	(70.6)
> 0.85	2,703	(28.7)	217	(29.4)
Age at menopause in years, mean (SD)	48	(6.20)	49	(6.26)
Total months of breastfeeding				
1-6	3,363	(35.8)	287	(38.9)*
7-12	4,451	(47.3)	357	(48.4)
> 13	1,591	(16.9)	93	(12.6)
Oral contraceptive duration in years, mean (SD)	6.71	(3.54)	5.16	(3.10)*
Oral contraceptive duration£				
< 5.1	2,745	(29.2)	381	(51.7)*
≥ 5.1	6,660	(70.8)	356	(48.3)
Exogenous estrogen use (E-only) in years				
Never	6,588	(70.0)	488	(66.2)*
< 5	1,312	(14.0)	106	(14.4)
5 to < 10	476	(5.1)	54	(7.3)
≥ 10	1,029	(10.9)	89	(12.1)
Exogenous estrogen use (E+P) in years				
Never	7,753	(82.4)	567	(76.9)*
< 5	900	(9.6)	86	(11.7)
5 to < 10	394	(4.2)	42	(5.7)
≥ 10	358	(3.8)	42	(5.7)

BMI, body mass index; CRC, colorectal cancer; E, estrogen; E+P, estrogen + progestin; MET, metabolic equivalent; MFA, monounsaturated fatty acids; PFA, polyunsaturated fatty acids; RSF, random survival forest; SFA, saturated fatty acids. *P < 0.05, chi-squared or Wilcoxon's rank-sum test. †Depression scales were estimated using a short form of the Center for Epidemiologic Studies Depression Scale. ¶Physical activity was estimated via recreational physical activity combining walking and mild, moderate, and strenuous physical activity. Each activity was assigned a MET value corresponding to intensity; the total MET·hours·week^{-1} was calculated by multiplying the MET level for the activity by the hours exercised per week and summing the values for all activities. The total MET was stratified into 2 groups, with 10 METs as the cutoff according to current American College of Sports Medicine and American Heart Association recommendations [97]. €Percent calories from SFA was classified by 9%, addressing low sample power (i.e., containing a quarter in one side) and adherent to the American Heart Association and American College of Cardiology dietary guidelines, which are aligned with the 2015-2020 Dietary Guidelines for Americans to help cardiovascular and metabolic diseases reductions [98]. ¥BMI, waist circumference, and waist-to-hip ratio were categorized at 30 kg/m^2, 88 cm, and 0.85, respectively, where those cutoff levels or higher fall within the overall or visceral obese range (https://www.cdc.gov/obesity/adult/defining.html; [99]). £Duration of oral contraceptive use was stratified at 5.1 years, where the cutoff level or higher fall within the high-risk group in the RSF model.

Table 2. The second stage of random survival forest analysis: predictive value of variables in overall analysis

Variable*	Minimal Depth†	VIMP	C-index	Error¶	Drop Error§
Duration of oral contraceptive use	1.9450	0.0467	0.8449	0.1551	0.3449
HNF4A rs1800961	2.5412	0.0263	0.9308	0.0692	0.0859
ONECUT2 rs4092465	2.6928	0.0179	0.9534	0.0466	0.0226
CRP rs1800947	2.9376	0.0115	0.9533	0.0467	-0.0001
METAP2 rs11108056	3.2086	0.0102	0.9534	0.0466	0.0001
NLRP3 rs10925027	3.4894	0.0073	0.9554	0.0446	0.0020
TOMM40 rs157581	3.9874	0.0119	0.9552	0.0448	-0.0002
TOMM40 rs157582	3.9934	0.0104	0.9562	0.0438	0.0011
TRAIP rs2352975	4.2272	0.0027	0.9580	0.0420	0.0017
DUSP1 rs17658229	4.2598	0.0047	0.9647	0.0353	0.0067
TOMM40 rs11556505	4.3526	0.0123	0.9657	0.0343	0.0011
Age at enrollment	4.5316	0.0015	0.9654	0.0346	-0.0003
Waist-to-hip ratio	5.0664	0.0000	0.9650	0.0350	-0.0005
RGS6 rs2239222	5.0824	0.0017	0.9642	0.0358	-0.0007
HNF1A rs11065385	5.0992	0.0041	0.9642	0.0358	0.0000
Duration of E+P use	5.1748	0.0009	0.9647	0.0353	0.0005
Age at menopause	5.3086	0.0003	0.9658	0.0342	0.0011
HNF1A-AS1 rs2251468	5.3192	0.0057	0.9652	0.0348	-0.0006
Hip circumference	5.3238	-0.0001	0.9647	0.0353	-0.0005
Height	5.4572	-0.0001	0.9641	0.0359	-0.0006
Education	5.5004	-0.0002	0.9634	0.0366	-0.0007
Waist circumference	5.5066	-0.0002	0.9629	0.0371	-0.0005
BMI	5.6638	-0.0001	0.9627	0.0373	-0.0002
Total months of breastfeeding	5.6970	0.0003	0.9621	0.0379	-0.0006
HNF1A-AS1 rs7953249	5.7660	0.0060	0.9618	0.0382	-0.0004
Weight	5.8404	-0.0001	0.9622	0.0378	0.0005
HNF1A-AS1 rs10774579	6.3998	0.0037	0.9623	0.0377	0.0001
HNF1A-AS1 rs1920792	6.4012	0.0037	0.9624	0.0376	0.0001
HNF1A rs1169301	8.0714	0.0010	0.9617	0.0383	-0.0007
HNF1A rs1169300	8.1956	0.0010	0.9616	0.0384	-0.0001

BMI, body mass index; C-index, concordance index; E+P, exogenous estrogen + progestin; VIMP, variable of importance. *Variables are ordered by minimal depth. †Predictive value of variable was assessed via minimal depth in the nested random survival forest models. A lower value is likely to have a greater impact on prediction. ¶The incremental error rate of each variable was estimated 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 estimated from the third nested model (including the first, second, and third variables). §The drop error rate was estimated by the difference between the error rates from the nested models with a prior and the corresponding variable. 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 (0.3449) from 0.5.

Combined and joint effects of the most predictive SNPs and lifestyle factor on CRC risk

Using the RSF model, we accounted for confounding factors and the nonlinearity of each variable and estimated the cumulative incidence rate of CRC (**Figure 3**). For the purpose of study analysis, we used this non-linear adjusted incidence rate to categorize the genotypes of each SNP that were originally continuous variables with the following risk genotypes (**Figure 3A-C**): *CRP* rs1800947 GG, *ONECUT2* rs4092465 GA, and *HNF4A* rs1800961 TT. Also, by using a cutoff-value diverging incidence

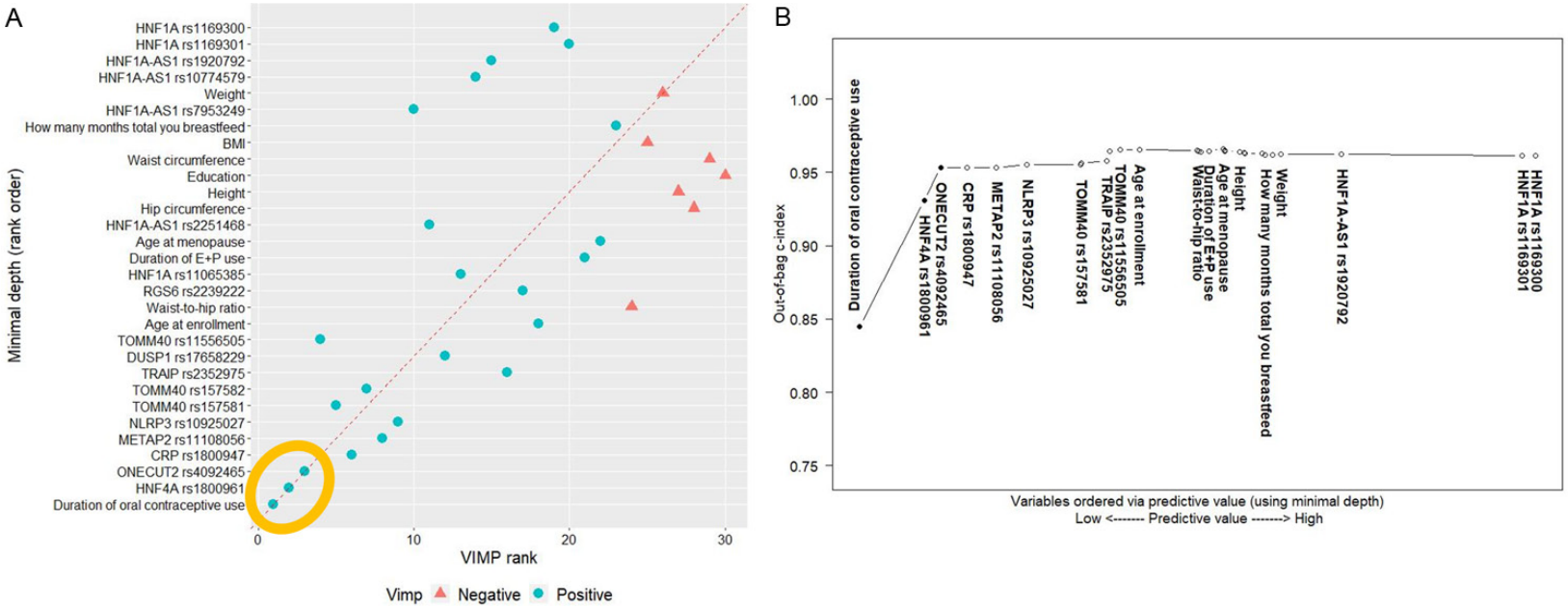


Figure 2. Overall analysis: the second stage of random survival forest (RSF) analysis with 18 single-nucleotide polymorphisms and 12 behavioral factors selected from the first stage of RSF. A. Comparison of minimal depth and VIMP rankings. (BMI, body mass index; E+P, exogenous estrogen + progestin; VIMP, variable of importance. Note: The 3 variables within the gold ellipse were identified as the most influential predictors. B. Out-of-bag concordance index (C-index). (Improvement in the out-of-bag C-index was observed when the top 3 variables [●] were added to the model, whereas other variables [○] did not further improve the accuracy of prediction).

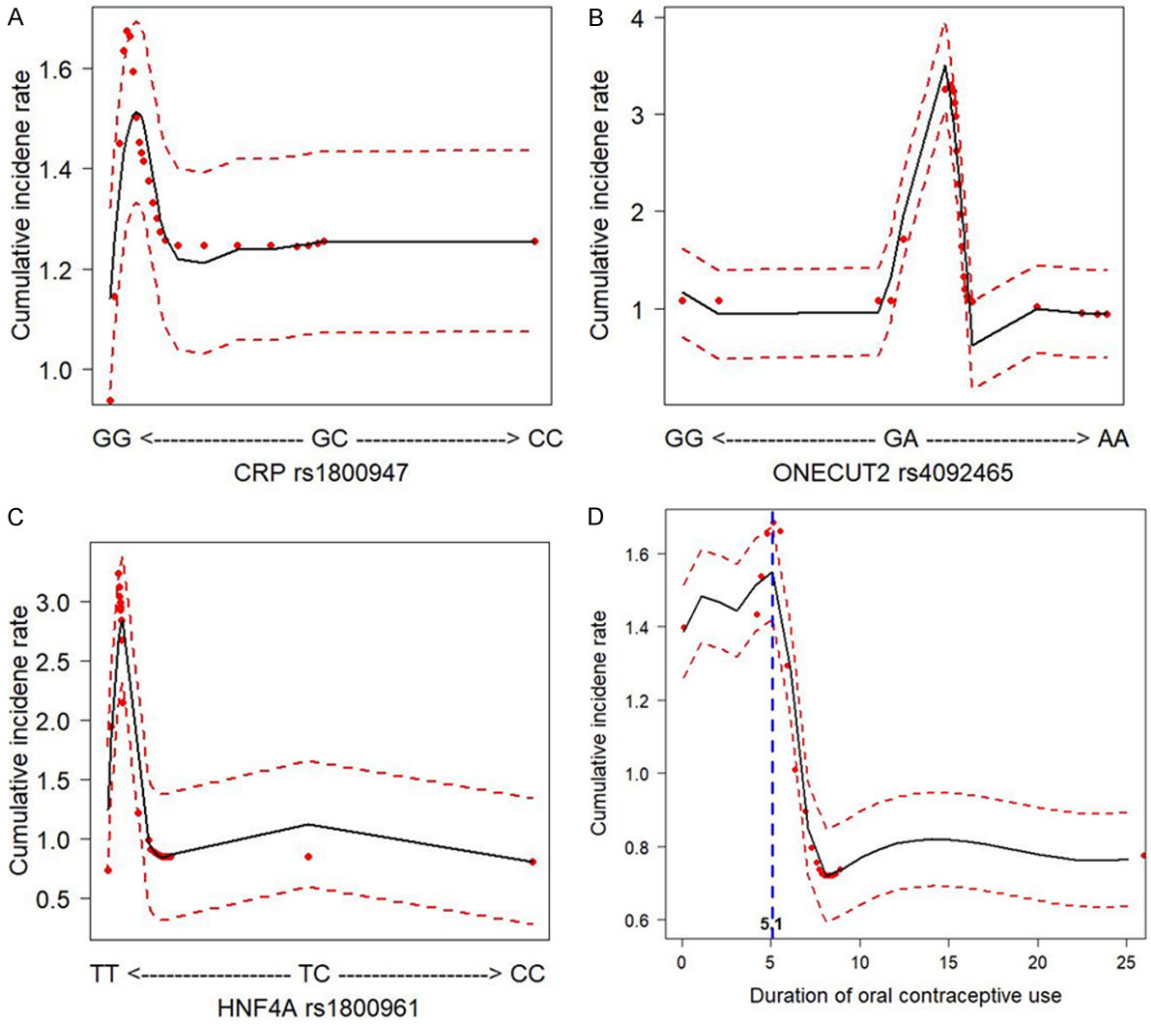


Figure 3. Cumulative colorectal cancer incidence rate for the 4 most influential variables (3 SNPs and 1 behavioral factor) based on random survival forest analyses. (SNPs, single-nucleotide polymorphisms. Dashed red lines indicate 95% confidence intervals).

rate of a variable (**Figure 3D**), we defined a high-risk lifestyle group as having < 5 years of past OC use, and further analyzed that as a binary variable. With the most predictive GMDR-modeled SNPs and risk lifestyle overall and in the obesity strata, we developed multivariate models predicting CRC risk (Tables S4A and S4B); the results indicated a stronger individual effect of 1 SNP (*ONECUT2* rs4092465 GA) than the other SNP and the risk lifestyle on CRC risk across the strata, even after adjusting for confounding factors.

The 2 top SNPs and the top risk lifestyle yielded different patterns when tested in combination or jointly for the association with CRC risk. For example, in the overall analysis, the top 2 SNPs were combined with the top risk lifestyle (< 5 years of past OC use) (**Table 3**). Compared with the lowest-risk group (i.e., low risk for genotypes and lifestyle), the moderate-risk (i.e., high risk of either genotypes or lifestyle) and the highest-risk groups (i.e., high risk of both genotypes and lifestyle) had about 4 times and 17 times the excessive risk, respectively, suggesting a gene-lifestyle dose-response relationship (**Table 3**). Next, we tested for the joint effect of past OC use with the risk genotypes of the 2 SNPs on CRC risk (**Table 4**). When the 2 SNPs were combined, their effect on CRC risk was not much different from the individual effect of *ONECUT2* rs4092465. However, when stratified by past OC use, the shorter-duration users (high risk: < 5.1 years) with the 2 risk alleles

Table 3. Combined effect of past OC use and risk genotypes (*ONECUT2* rs4092465 GA; *HNF4A* rs1800961 TT) on CRC risk overall and in BMI strata

n£	Risk genotype + past OC use		
	Total n	HR† (95% CI)	*p**
	<Overall group>		
0	3,559	reference	
1	4,949	**4.40 (3.32-5.83)**	**< 2e-16**
2	1,634	**17.54 (13.22-23.28)**	**< 2e-16**
p_{trend}			< 2e-16
	<Non-overall obese group, BMI < 30 kg/m^2 (n=7,151)>		
0	2,594	reference	
1	3,501	**4.23 (3.08-5.82)**	**< 2e-16**
2	1,056	**17.41 (12.62-24.02)**	**< 2e-16**
p_{trend}			< 2e-16
	<Overall obese group, BMI ≥ 30 kg/m^2 (n=2,991)>		
0	965	reference	
1	1,448	**5.08 (2.77-9.30)**	**1.37e-07**
2	578	**19.96 (10.91-36.51)**	**< 2e-16**
p_{trend}			< 2e-16

BMI, body mass index; CI, confidence interval; CRC, colorectal cancer; HR, hazard ratio; OC, oral contraceptive. Numbers in bold face are statistically significant. £The combined number of risk genotypes and behavioral factors was based on 1) risk genotypes defined as 0 (low risk: none or 1 risk allele) and 1 (high risk: 2 risk alleles) and 2) behavioral factors defined as 0 (low risk: past OC use ≥ 5.1 years) and 1 (high risk: past OC use < 5.1 years). The ultimate number of combined risk genotypes and behavioral factors was defined as 0 (low risk for genotypes and behaviors), 1 (high risk for either genotypes or behaviors), and 2 (high risk for both genotypes and behaviors). †Multivariate regression was adjusted by age at enrollment, education, BMI (in overall group), height and weight (in BMI strata), waist-to-hip ratio, age at menopause, total months of breastfeeding, and exogenous estrogen plus progestin. **p* values were adjusted to correct for multiple testing via the Benjamini-Hochberg approach.

had 3 times higher risk for CRC than the longer-duration users (low risk: ≥ 5 years) with the same 2 risk alleles, and further, 17 times greater risk than the longer-duration users with null or 1 risk allele. This indicates a significant joint effect of past OC use with the risk genotypes on CRC risk in both additive and multiplicative interaction models (G×E: HR=1.98, P=0.003). Multiple testing corrections were made to control the false-discovery rate. In the BMI strata (**Tables 3** and **4**), the non-overall obese group yielded similar results to those from the overall analysis, but the overall-obese group displayed stronger combined and joint effects of the risk genotypes and past OC use in both additive and multiplicative models (G×E: HR=1.54, P=0.332); these results suggest the potential existence of the BMI-modified inflammatory gene-lifestyle pathway predicting CRC risk.

Further, we examined the combined and joint effects of past OC use with the risk genotypes on CRC risk in the other obesity strata, and determined that the risk genotypes in combination (Table S5) or jointly (Table S6) associated with OC use displayed a synergistic effect on CRC risk; particularly in the METs strata, the gene-lifestyle combined and joint effect on cancer appeared more strongly in the physically inactive group (**Figure 4**).

Discussion

An increasing number of population-based human genomic studies have incorporated environmental factors in cancer molecular causal pathways. Elucidating the role of lifestyle factors in modifying the gene and phenotype association, thus influencing the risk of CRC, may improve the predictability for CRC and facilitate the development of personalized genetically targeted lifestyle interventions for primary cancer prevention efforts. Our multimodal RSF and GMDR approaches determined the best genetic markers in inflammatory cytokines and the best risk lifestyle predictive for CRC development. The most common predictors across the obesity strata are 2 SNPs (*ONECUT2* rs4092465 and *HNF4A* rs1800961) and 1 lifestyle factor (relatively short-term past use of OC). The risk profiles that combined those genetic and lifestyle markers exhibited a synergistically increased risk for CRC, and this pattern appeared more strongly in obese subgroups.

ONECUT2 encodes the second member of the ONECUT family of DNA-binding transcription factors, characterized by a single cut domain and a specific homeodomain, and its phenotypes have been linked to CRP and stroke [72, 73]. ONECUT2, as an angiogenic and epithelial-

Table 4. Joint effect of past OC use with risk genotypes (*ONECUT2* rs4092465 GA; *HNF4A* rs1800961 TT) on CRC risk overall and in BMI strata

n	Total			Past OC use ≥ 5.1 years			Past OC use < 5.1 years		
	HR† (95% CI)	p*	n	HR† (95% CI)	*p**	n	HR† (95% CI)	*p**
<Overall group>								
Risk genotypes£								
0	reference		3,559	reference		1,492	**1.59 (1.04-2.43)**	**0.03415**
1	**7.39 (5.93-9.20)**	**< 2e-16**	3,457	**5.46 (4.11-7.25)**	**< 2e-16**	1,634	**17.15 (12.92-22.76)**	**< 2e-16**
			p_{trend}		< 2e-16			
<Non-overall obese group, BMI < 30 kg/m^2 (n=7,151)>								
Risk genotypes£								
0	reference		2,594	reference		946	1.52 (0.92-2.53)	0.10148
1	**7.07 (5.48-9.12)**	**< 2e-16**	2,555	**5.14 (3.73-7.07)**	**< 2e-16**	1,056	**17.05 (12.36-23.53)**	**< 2e-16**
			p_{trend}		< 2e-16			
<Overall obese group, BMI ≥ 30 kg/m^2 (n=2,991)>								
Risk genotypes£								
0	reference		965	reference		546	1.89 (0.83-4.30)	0.12766
1	**8.40 (5.43-12.97)**	**< 2e-16**	902	**6.69 (3.64-12.31)**	**1e-09**	578	**19.50 (10.66-35.68)**	**< 2e-16**
			p_{trend}		< 2e-16			

BMI, body mass index; CI, confidence interval; CRC, colorectal cancer; HR, hazard ratio; OC, oral contraceptive. Numbers in bold face are statistically significant. †Multivariate regression was adjusted by age at enrollment, education, BMI (in overall group), height and weight (in BMI strata), waist-to-hip ratio, age at menopause, total months of breastfeeding, OC use (in total analysis), and exogenous estrogen plus progestin. **p* values were adjusted to correct for multiple testing via the Benjamini-Hochberg approach. £The number of risk genotypes was defined on the basis of Kaplan-Meier analysis as follows: 0 (none or 1 risk allele) and 1 (2 risk alleles).

mesenchymal transition (EMT) marker, plays a key role in the oncogenesis of several cancers, including hepatocellular carcinoma (HCC) [74] and ovarian [73], breast [75], and prostate cancers [76]. With regard to CRC, ONECUT2 is involved in the EMT and the migration and invasion of CRC cells, and it acts as a tumor promoter [77], although the colorectal carcinogenic mechanism is not fully determined. Our genomic study initially detected 1 SNP near *ONECUT2* for the association with CRP at the GWA level and its strong association with CRC risk, suggesting the effect of the genetic aberration in this transcription factor on colorectal tumorigenesis. Notably, the effect of this SNP on CRC risk appeared much stronger when it was combined and jointly interacted with the cumulative exposure to estrogen, warranting future biologic mechanism study on the inflammatory-sex hormone interaction in colorectal carcinogenesis.

HNF4A encodes hepatocyte nuclear factor 4 alpha, a nuclear transcription factor that controls the expression of other genes such as *HNF1A*, and it plays a role in the maturation of liver, kidney, and intestine [78, 79]. Its phenotypes include CRP [32, 45, 80, 81] and high-density lipoprotein [82], and mutations in this gene have been associated with maturity-onset diabetes of the young type 1, type 2 diabetes, ulcerative colitis, and Crohn's disease [82, 83]. *HNF4A* expression has antitumor activity in several cancer cell lines, such as lung carcinoma [84, 85], head and neck squamous cell carcinoma [86], esophageal adenocarcinoma [87], and HCC [88, 89]. Also, the ectopic expression of *HNF4A* inhibited CRC cells' proliferation, invasion, and migration by G2/M cell-phase arrest and promoted apoptosis through its effect on the Wnt/B-catenin signaling pathway [90]. Further, HNF4A protects the intestinal mucosa against inflammation by blocking the *IL6R/STAT3* pathway [83]. Therefore, downregulation of *HNF4A* expression is crucial in the aggravation of CRC. As supported by these previous studies, the minor allele of the *HNF4A* SNP in our study, both individually and in combination with the *ONECUT2* SNP, presented a strong effect on increased risk for CRC. Further, considering the positive correlation between the *HNF4A* genetic score and BMI [91], our finding of the greater impact of the SNPs on CRC risk in the BMI-obese and MET-inactive subgroups is biologically compelling. Additionally, among obese or diabetic individuals (i.e., those

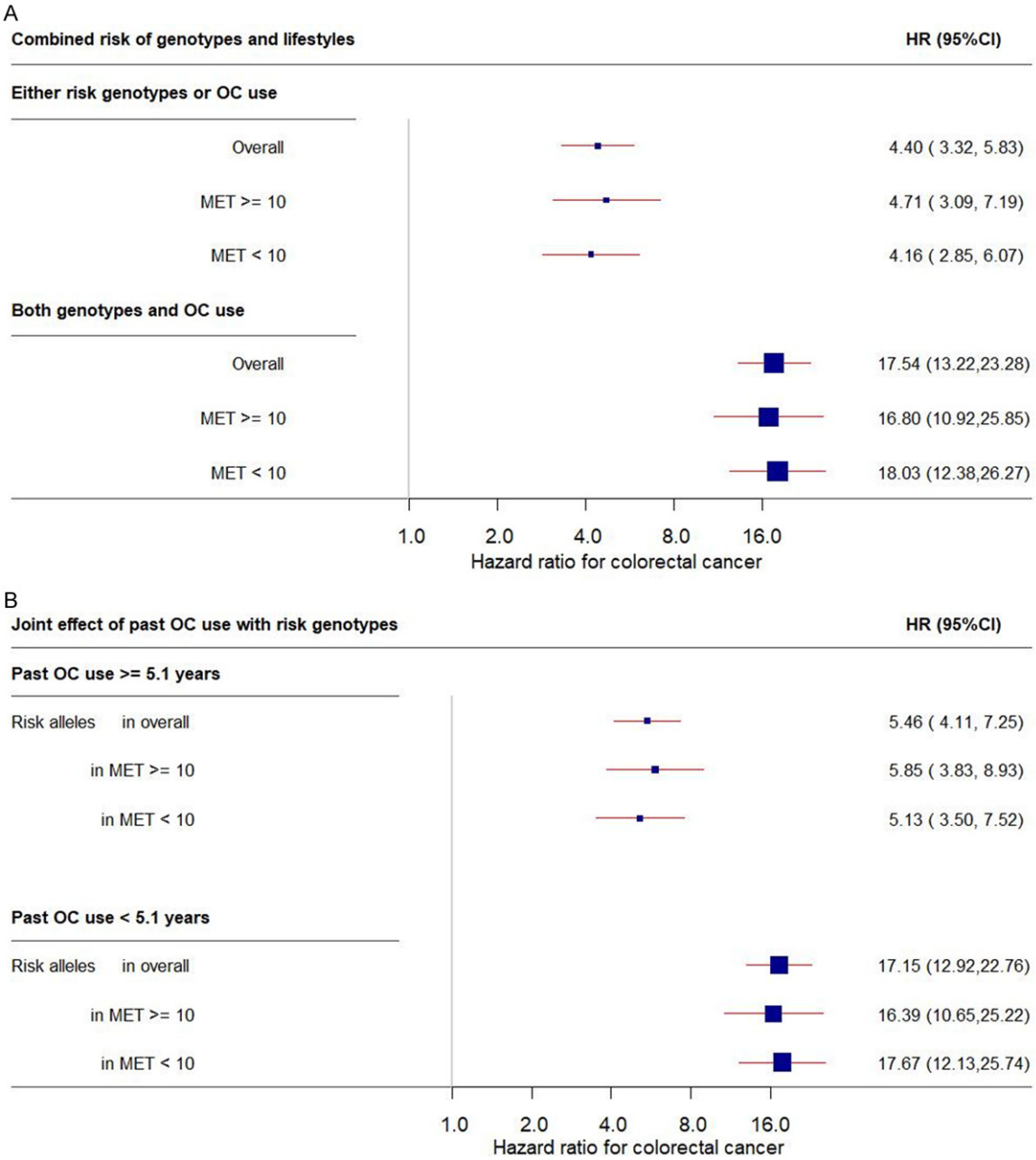


Figure 4. Forest plot of the combined (A) and joint (B) effect of past OC use and risk genotypes on CRC risk overall and in MET subgroups. Plot (A) shows the independent and combined effect of risk genotypes and OC use on CRC risk, and Plot (B) shows the joint tests for risk genotypes with OC use, presented as the 95% CIs (indicated with red lines) and the estimates (proportional to the size of the blue squares). The analyzed risk genotypes included *ONECUT2* rs4092465 GA and *HNF4A* rs1800961 TT. (CI, confidence interval; CRC, colorectal cancer; HR, hazard ratio; MET, metabolic equivalent; OC, oral contraceptive).

with chronic low-grade inflammatory diseases), HNF4A has hepatic cross-talk with sex hormones such as sex hormone-binding globulin [92, 93], potentially influencing cancer initiation and progress, although the underlying mechanism is uncertain. Our study indicates that the *HNF4A/ONECUT2* SNPs combined with past OC use synergistically increase CRC risk, and that this pattern is more profound in obese subgroups; this may provide a scientific basis for future mechanistic study.

Lifetime cumulative exposure to estrogen, particularly past use of OC in postmenopausal

women, has been considered a protective factor for CRC risk, as shown in *in vivo* and *in vitro* studies reporting that estrogen upregulates a p53 cell-cycle regulator, leading to the inhibition of CRC cell growth [94, 95]. Our RSF-based estimation of the CRC cumulative incidence rate for OC use accounted for its nonlinearity, showing that cancer risk increased with up to 5 years of OC use but dropped thereafter. The past use of OC, when combined with inflammatory SNPs, increased cancer risk in a gene-lifestyle dose response-associated manner, implying the existence of inflammatory-sex hormone cross-talk in colorectal carcinogenesis.

In the early 1980s, OC formulations had high estrogen concentration, but that has since been changed [96]. We did not have data available on OC types. Because CRC risk can be dependent on different OC preparations, our study should be further validated with data incorporating information on OC estrogen concentration. We also had no data on CRC molecular subtypes, warranting future independent studies with data on the tumor molecular features. Further, our study population was restricted to non-Hispanic white postmenopausal women, so our findings should not be generalized to other populations. Despite the strong benefits from the 2-stage multimodal approaches, they are noisy tasks, leading to the over-fit model in small subgroups; thus, a replication study with a large sample size is requested.

In this study, the GWA SNPs in pro-inflammatory cytokines exhibit synergistically increased risk for CRC in combination with lifetime cumulative exposure to estrogen and this pattern is more profound in obese/inactive subgroups. Our results call for molecular studies to examine inflammatory gene signatures and their expressions aberrantly presenting in their downstream signaling pathways in relation to CRC by incorporating the effect of estrogen and obesity. Our study may contribute to an improved prediction ability for CRC risk and suggest genetically targeted lifestyle interventions for women carrying the inflammatory-risk genotypes, thus reducing CRC risk.

Acknowledgements

This study was supported by the National Institute of Nursing Research of the National Institutes of Health under Award Number K01NR017852 and a University of California Cancer Research Coordinating Committee grant (CRN-18-522722). The data for this project were 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).

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Su Yon Jung, Translational Sciences Section, Jonsson Comprehensive Cancer Center, School of Nursing, University of California, 700 Tiverton Ave, 3-264 Factor Building, Los Angeles, CA 90095, USA. Tel: 310-825-2840; Fax: 310-267-0413; E-mail: sjung@sonnet.ucla.edu

References

[1] Coussens LM and Werb Z. Inflammation and cancer. Nature 2002; 420: 860-867.
[2] Disis ML. Immune regulation of cancer. J Clin Oncol 2010; 28: 4531-4538.
[3] American Cancer Society. Colorectal Cancer Facts & Figures 2017-2019. Atlanta: American Cancer Society, Inc. 2017: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/colorectal-cancer-facts-and-figures/colorectal-cancer-facts-and-figures-2017-2019.pdf.
[4] American Cancer Society. Cancer Fact and Figures, 2020. Atlanta: American Cancer Society, Inc.: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf.
[5] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018; 68: 394-424.
[6] Pohl C, Hombach A and Kruis W. Chronic inflammatory bowel disease and cancer. Hepatogastroenterology 2000; 47: 57-70.
[7] Erlinger TP, Platz EA, Rifai N and Helzlsouer KJ. C-reactive protein and the risk of incident colorectal cancer. JAMA 2004; 291: 585-590.

[8] Prizment AE, Anderson KE, Visvanathan K and Folsom AR. Association of inflammatory markers with colorectal cancer incidence in the atherosclerosis risk in communities study. Cancer Epidemiol Biomarkers Prev 2011; 20: 297-307.

[9] Toriola AT, Cheng TY, Neuhouser ML, Wener MH, Zheng Y, Brown E, Miller JW, Song X, Beresford SA, Gunter MJ, Caudill MA and Ulrich CM. Biomarkers of inflammation are associated with colorectal cancer risk in women but are not suitable as early detection markers. Int J Cancer 2013; 132: 2648-2658.

[10] Ho GY, Wang T, Gunter MJ, Strickler HD, Cushman M, Kaplan RC, Wassertheil-Smoller S, Xue X, Rajpathak SN, Chlebowski RT, Vitolins MZ, Scherer PE and Rohan TE. Adipokines linking obesity with colorectal cancer risk in postmenopausal women. Cancer Res 2012; 72: 3029-3037.

[11] Ho GY, Wang T, Zheng SL, Tinker L, Xu J, Rohan TE, Wassertheil-Smoller S, Xue X, Augenlicht LH, Peters U, Phipps AI, Strickler HD, Gunter MJ and Cushman M. Circulating soluble cytokine receptors and colorectal cancer risk. Cancer Epidemiol Biomarkers Prev 2014; 23: 179-88.

[12] Ghazi-Khanloosani M, Bandegi AR, Kokhaei P, Barati M and Pakdel A. CRP and LOX-1: a mechanism for increasing the tumorigenic potential of colorectal cancer carcinoma cell line. Pathol Oncol Res 2019; 25: 1467-1475.

[13] Cui G, Yuan A, Sun Z, Zheng W and Pang Z. IL-1beta/IL-6 network in the tumor microenvironment of human colorectal cancer. Pathol Res Pract 2018; 214: 986-992.

[14] Liu L, Tabung FK, Zhang X, Nowak JA, Qian ZR, Hamada T, Nevo D, Bullman S, Mima K, Kosumi K, da Silva A, Song M, Cao Y, Twombly TS, Shi Y, Liu H, Gu M, Koh H, Li W, Du C, Chen Y, Li C, Li W, Mehta RS, Wu K, Wang M, Kostic AD, Giannakis M, Garrett WS, Hutthenhower C, Chan AT, Fuchs CS, Nishihara R, Ogino S and Giovannucci EL. Diets that promote colon inflammation associate with risk of colorectal carcinomas that contain fusobacterium nucleatum. Clin Gastroenterol Hepatol 2018; 16: 1622-1631, e3.

[15] Zhang Y, Yu X, Yu E, Wang N, Cai Q, Shuai Q, Yan F, Jiang L, Wang H, Liu J, Chen Y, Li Z and Jiang Q. Changes in gut microbiota and plasma inflammatory factors across the stages of colorectal tumorigenesis: a case-control study. BMC Microbiol 2018; 18: 92.

[16] Proença MA, Biselli JM, Succi M, Severino FE, Berardinelli GN, Caetano A, Reis RM, Hughes DJ and Silva AE. Relationship between Fusobacterium nucleatum, inflammatory mediators and microRNAs in colorectal carcinogenesis. World J Gastroenterol 2018; 24: 5351-5365.

[17] Castellano-Castillo D, Morcillo S, Clemente-Postigo M, Crujeiras AB, Fernandez-Garcia JC, Torres E, Tinahones FJ and Macias-Gonzalez M. Adipose tissue inflammation and VDR expression and methylation in colorectal cancer. Clin Epigenetics 2018; 10: 60.

[18] Himbert C, Ose J, Nattenmüller J, Warby CA, Holowatyj AN, Böhm J, Lin T, Haffa M, Gigic B, Hardikar S, Scherer D, Zielske L, Schrotz-King P, Kölsch T, Siegel EM, Shibata D, Ulrich A, Schneider M, Hursting SD, Kauczor HU and Ulrich CM. Body fatness, adipose tissue compartments, and biomarkers of inflammation and angiogenesis in colorectal cancer: the colocare study. Cancer Epidemiol Biomarkers Prev 2019; 28: 76-82.

[19] Pfalzer AC, Leung K, Crott JW, Kim SJ, Tai AK, Parnell LD, Kamanu FK, Liu Z, Rogers G, Shea MK, Garcia PE and Mason JB. Incremental elevations in TNFalpha and IL6 in the human colon and procancerous changes in the mucosal transcriptome accompany adiposity. Cancer Epidemiol Biomarkers Prev 2018; 27: 1416-1423.

[20] Fang D and Ye Y. C-reactive protein gene rs1205 polymorphism is not associated with the risk of colorectal cancer. Biosci Rep 2017; 37: BSR20170872.

[21] Huang BZ, Tsilidis KK, Smith MW, Hoffman-Bolton J, Visvanathan K, Platz EA and Joshu CE. Polymorphisms in genes related to inflammation and obesity and colorectal adenoma risk. Mol Carcinog 2018; 57: 1278-1288.

[22] Hu JJ, Wang ZT and Zhong J. Lack of association between the interleukin 6 gene -174G>C polymorphism and colorectal cancer: evidence from a meta-analysis. Genet Mol Res 2013; 12: 2205-14.

[23] Banday MZ, Balkhi HM, Sameer AS, Chowdri NA and Haq E. Strong association of interleukin-6 -174G/C promoter single nucleotide polymorphism with a decreased risk of colorectal cancer in ethnic Kashmiri population: a case control study. Tumour Biol 2017; 39: 1010428317695940.

[24] Landi S, Moreno V, Gioia-Patricola L, Guino E, Navarro M, de Oca J, Capella G and Canzian F; Bellvitge Colorectal Cancer Study Group. Association of common polymorphisms in inflammatory genes interleukin (IL)6, IL8, tumor necrosis factor alpha, NFKB1, and peroxisome proliferator-activated receptor gamma with colorectal cancer. Cancer Res 2003; 63: 3560-3566.

[25] Wang X, Dai JY, Albanes D, Arndt V, Berndt SI, Bézieau S, Brenner H, Buchanan DD, Butterbach K, Caan B, Casey G, Campbell PT,

Chan AT, Chen Z, Chang-Claude J, Cotterchio M, Easton DF, Giles GG, Giovannucci E, Grady WM, Hoffmeister M, Hopper JL, Hsu L, Jenkins MA, Joshi AD, Lampe JW, Larsson SC, Lejbkowicz F, Li L, Lindblom A, Le Marchand L, Martin V, Milne RL, Moreno V, Newcomb PA, Offitt K, Ogino S, Pharoah PDP, Pinchev M, Potter JD, Rennert HS, Rennert G, Saliba W, Schafmayer C, Schoen RE, Schrotz-King P, Slattery ML, Song M, Stegmaier C, Weinstein SJ, Wolk A, Woods MO, Wu AH, Gruber SB, Peters U and White E. Mendelian randomization analysis of C-reactive protein on colorectal cancer risk. Int J Epidemiol 2019; 48: 767-780.

[26] Amaral WZ, Krueger RF, Ryff CD and Coe CL. Genetic and environmental determinants of population variation in interleukin-6, its soluble receptor and C-reactive protein: insights from identical and fraternal twins. Brain Behav Immun 2015; 49: 171-181.

[27] Fried SK, Bunkin DA and Greenberg AS. Omental and subcutaneous adipose tissues of obese subjects release interleukin-6: depot difference and regulation by glucocorticoid. J Clin Endocrinol Metab 1998; 83: 847-850.

[28] Ligthart S, Vaez A, Võsa U, Stathopoulou MG, Ligthart S, Vaez A, Võsa U, Stathopoulou MG, de Vries PS, Prins BP, Van der Most PJ, Tanaka T, Naderi E, Rose LM, Wu Y, Karlsson R, Barbalic M, Lin H, Pool R, Zhu G, Macé A, Sidore C, Trompet S, Mangino M, Sabater-Lleal M, Kemp JP, Abbasi A, Kacprowski T, Verweij N, Smith AV, Huang T, Marzi C, Feitosa MF, Lohman KK, Kleber ME, Milaneschi Y, Mueller C, Huq M, Vlachopoulou E, Lyytikäinen LP, Oldmeadow C, Deelen J, Perola M, Zhao JH, Feenstra B; LifeLines Cohort Study, Amini M; CHARGE Inflammation Working Group, Lahti J, Schraut KE, Fornage M, Suktitipat B, Chen WM, Li X, Nutile T, Malerba G, Luan J, Bak T, Schork N, Del Greco M F, Thiering E, Mahajan A, Marioni RE, Mihailov E, Eriksson J, Ozel AB, Zhang W, Nethander M, Cheng YC, Aslibekyan S, Ang W, Gandin I, Yengo L, Portas L, Kooperberg C, Hofer E, Rajan KB, Schurmann C, den Hollander W, Ahluwalia TS, Zhao J, Draisma HHM, Ford I, Timpson N, Teumer A, Huang H, Wahl S, Liu Y, Huang J, Uh HW, Geller F, Joshi PK, Yanek LR, Trabetti E, Lehne B, Vozzi D, Verbanck M, Biino G, Saba Y, Meulenbelt I, O'Connell JR, Laakso M, Giulianini F, Magnusson PKE, Ballantyne CM, Hottenga JJ, Montgomery GW, Rivadineira F, Rueedi R, Steri M, Herzig KH, Stott DJ, Menni C, Frånberg M, St Pourcain B, Felix SB, Pers TH, Bakker SJL, Kraft P, Peters A, Vaidya D, Delgado G, Smit JH, Großmann V, Sinisalo J, Seppälä I, Williams SR, Holliday EG, Moed M, Langenberg C, Räikkönen K, Ding J, Campbell H, Sale MM, Chen YI, James AL, Ruggiero D, Soranzo N, Hartman CA, Smith EN, Berenson GS, Fuchsberger C, Hernandez D, Tiesler CMT, Giedraitis V, Liewald D, Fischer K, Mellström D, Larsson A, Wang Y, Scott WR, Lorentzon M, Beilby J, Ryan KA, Pennell CE, Vuckovic D, Balkau B, Concas MP, Schmidt R, Mendes de Leon CF, Bottinger EP, Kloppenburg M, Paternoster L, Boehnke M, Musk AW, Willemsen G, Evans DM, Madden PAF, Kähönen M, Kutalik Z, Zoledziewska M, Karhunen V, Kritchevsky SB, Sattar N, Lachance G, Clarke R, Harris TB, Raitakari OT, Attia JR, van Heemst D, Kajantie E, Sorice R, Gambaro G, Scott RA, Hicks AA, Ferrucci L, Standl M, Lindgren CM, Starr JM, Karlsson M, Lind L, Li JZ, Chambers JC, Mori TA, de Geus EJCN, Heath AC, Martin NG, Auvinen J, Buckley BM, de Craen AJM, Waldenberger M, Strauch K, Meitinger T, Scott RJ, McEvoy M, Beekman M, Bombieri C, Ridker PM, Mohlke KL, Pedersen NL, Morrison AC, Boomsma DI, Whitfield JB, Strachan DP, Hofman A, Vollenweider P, Cucca F, Jarvelin MR, Jukema JW, Spector TD, Hamsten A, Zeller T, Uitterlinden AG, Nauck M, Gudnason V, Qi L, Grallert H, Borecki IB, Rotter JI, März W, Wild PS, Lokki ML, Boyle M, Salomaa V, Melbye M, Eriksson JG, Wilson JF, Penninx BWJH, Becker DM, Worrall BB, Gibson G, Krauss RM, Ciullo M, Zaza G, Wareham NJ, Oldehinkel AJ, Palmer LJ, Murray SS, Pramstaller PP, Bandinelli S, Heinrich J, Ingelsson E, Deary IJ, Mägi R, Vandenput L, van der Harst P, Desch KC, Kooner JS, Ohlsson C, Hayward C, Lehtimäki T, Shuldiner AR, Arnett DK, Beilin LJ, Robino A, Froguel P, Pirastu M, Jess T, Koenig W, Loos RJF, Evans DA, Schmidt H, Smith GD, Slagboom PE, Eiriksdottir G, Morris AP, Psaty BM, Tracy RP, Nolte IM, Boerwinkle E, Visvikis-Siest S, Reiner AP, Gross M, Bis JC, Franke L, Franco OH, Benjamin EJ, Chasman DI, Dupuis J, Snieder H, Dehghan A and Alizadeh BZ. Genome analyses of >200,000 individuals identify 58 loci for chronic inflammation and highlight pathways that link inflammation and complex disorders. Am J Hum Genet 2018; 103: 691-706.

[29] Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, Pellikka N, Wallaschofski H, Kettunen J, Henneman P, Baumert J, Strachan DP, Fuchsberger C, Vitart V, Wilson JF, Paré G, Naitza S, Rudock ME, Surakka I, de Geus EJ, Alizadeh BZ, Guralnik J, Shuldiner A, Tanaka T, Zee RY, Schnabel RB, Nambi V, Kavousi M, Ripatti S, Nauck M, Smith NL, Smith AV, Sundvall J, Scheet P, Liu Y, Ruokonen A, Rose LM, Larson MG, Hoogeveen RC, Freimer NB, Teumer A, Tracy RP, Launer LJ, Buring JE, Yamamoto JF, Folsom AR, Sijbrands

EJ, Pankow J, Elliott P, Keaney JF, Sun W, Sarin AP, Fontes JD, Badola S, Astor BC, Hofman A, Pouta A, Werdan K, Greiser KH, Kuss O, Meyer zu Schwabedissen HE, Thiery J, Jamshidi Y, Nolte IM, Soranzo N, Spector TD, Völzke H, Parker AN, Aspelund T, Bates D, Young L, Tsui K, Siscovick DS, Guo X, Rotter JI, Uda M, Schlessinger D, Rudan I, Hicks AA, Penninx BW, Thorand B, Gieger C, Coresh J, Willemsen G, Harris TB, Uitterlinden AG, Järvelin MR, Rice K, Radke D, Salomaa V, Willems van Dijk K, Boerwinkle E, Vasan RS, Ferrucci L, Gibson QD, Bandinelli S, Snieder H, Boomsma DI, Xiao X, Campbell H, Hayward C, Pramstaller PP, van Duijn CM, Peltonen L, Psaty BM, Gudnason V, Ridker PM, Homuth G, Koenig W, Ballantyne CM, Witteman JC, Benjamin EJ, Perola M and Chasman DI. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. Circulation 2011; 123: 731-738.

[30] Doumatey AP, Chen G, Tekola Ayele F, Zhou J, Erdos M, Shriner D, Huang H, Adeleye J, Balogun W, Fasanmade O, Johnson T, Oli J, Okafor G, Amoah A, Eghan BA, Agyenim-Boateng K, Acheampong J, Adebamowo C, Gerry NP, Christman MF, Adeyemo A and Rotimi CN. C-reactive protein (CRP) promoter polymorphisms influence circulating CRP levels in a genome-wide association study of African Americans. Hum Mol Genet 2012; 21: 3063-3072.

[31] Ridker PM, Pare G, Parker A, Zee RY, Danik JS, Buring JE, Kwiatkowski D, Cook NR, Miletich JP and Chasman DI. Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: the women's genome health study. Am J Hum Genet 2008; 82: 1185-1192.

[32] Reiner AP, Beleza S, Franceschini N, Auer PL, Robinson JG, Kooperberg C, Peters U and Tang H. Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. Am J Hum Genet 2012; 91: 502-512.

[33] Prizment AE, Folsom AR, Dreyfus J, Anderson KE, Visvanathan K, Joshu CE, Platz EA and Pankow JS. Plasma C-reactive protein, genetic risk score, and risk of common cancers in the atherosclerosis risk in communities study. Cancer Causes Control 2013; 24: 2077-2087.

[34] Hu M, Lee MH, Mak VW and Tomlinson B. Effect of central obesity, low high-density lipoprotein cholesterol and C-reactive protein polymorphisms on C-reactive protein levels during treatment with Rosuvastatin (10 mg Daily). Am J Cardiol 2010; 106: 1588-1593.

[35] Wu S, Hsu LA, Teng MS, Lin JF, Chou HH, Lee MC, Wu YM, Su CW and Ko YL. Interactive effects of C-reactive protein levels on the association between APOE variants and triglyceride levels in a Taiwanese population. Lipids Health Dis 2016; 15: 94.

[36] Fraser A, May M, Lowe G, Rumley A, Smith GD, Ebrahim S and Lawlor DA. Interleukin-6 and incident coronary heart disease: results from the British Women's Heart and Health Study. Atherosclerosis 2009; 202: 567-572.

[37] Winters-Stone KM, Wood LJ, Stoyles S and Dieckmann NF. The effects of resistance exercise on biomarkers of breast cancer prognosis: a pooled analysis of three randomized trials. Cancer Epidemiol Biomarkers Prev 2018; 27: 146-153.

[38] Lynch BM, Friedenreich CM, Winkler EA, Healy GN, Vallance JK, Eakin EG and Owen N. Associations of objectively assessed physical activity and sedentary time with biomarkers of breast cancer risk in postmenopausal women: findings from NHANES (2003-2006). Breast Cancer Res Treat 2011; 130: 183-194.

[39] van Gemert WA, May AM, Schuit AJ, Oosterhof BY, Peeters PH and Monninkhof EM. Effect of weight loss with or without exercise on inflammatory markers and adipokines in postmenopausal women: the SHAPE-2 trial, a randomized controlled trial. BMC Public Health 2016; 25: 799-806.

[40] Rojo-Martínez G, Soriguer F, Colomo N, Calle A, Goday A, Bordiú E, Delgado E, Menéndez E, Ortega E, Urrutia I, Girbés J, Castaño L, Catalá M, Gaztambide S and Valdés S; di@bet.es study group.. Factors determining high-sensitivity C-reactive protein values in the Spanish population. Di@bet.es study. Eur J Clin Invest 2013; 43: 1-10.

[41] Dias JA, Wirfalt E, Drake I, Gullberg B, Hedblad B, Persson M, Engström G, Nilsson J, Schiopu A, Fredrikson GN and Björkbacka H. A high quality diet is associated with reduced systemic inflammation in middle-aged individuals. Atherosclerosis 2015; 238: 38-44.

[42] Bermudez EA, Rifai N, Buring JE, Manson JE and Ridker PM. Relation between markers of systemic vascular inflammation and smoking in women. Am J Cardiol 2002; 89: 1117-1119.

[43] Stewart SH, Mainous AG 3rd and Gilbert G. Relation between alcohol consumption and C-reactive protein levels in the adult US population. J Am Board Fam Pract 2002; 15: 437-42.

[44] Jung SY, Scott P, Papp J, Sobel E, Pellegrini M, Yu H, et al. Genome-wide association analysis of pro-inflammatory cytokines and gene-lifestyle interaction for invasive breast cancer risk: the WHI dbGaP Study. Submitted and Reviewed for Publication. 2020.

[45] Ellis J, Lange EM, Li J, Dupuis J, Baumert J, Walston JD, Keating BJ, Durda P, Fox ER,

Palmer CD, Meng YA, Young T, Farlow DN, Schnabel RB, Marzi CS, Larkin E, Martin LW, Bis JC, Auer P, Ramachandran VS, Gabriel SB, Willis MS, Pankow JS, Papanicolaou GJ, Rotter JI, Ballantyne CM, Gross MD, Lettre G, Wilson JG, Peters U, Koenig W, Tracy RP, Redline S, Reiner AP, Benjamin EJ and Lange LA. Large multiethnic Candidate Gene Study for C-reactive protein levels: identification of a novel association at CD36 in African Americans. Hum Genet 2014; 133: 985-995.

[46] Schick UM, Auer PL, Bis JC, Lin H, Wei P, Pankratz N, Lange LA, Brody J, Stitziel NO, Kim DS, Carlson CS, Fornage M, Haessler J, Hsu L, Jackson RD, Kooperberg C, Leal SM, Psaty BM, Boerwinkle E, Tracy R, Ardissino D, Shah S, Willer C, Loos R, Melander O, Mcpherson R, Hovingh K, Reilly M, Watkins H, Girelli D, Fontanillas P, Chasman DI, Gabriel SB, Gibbs R, Nickerson DA, Kathiresan S, Peters U, Dupuis J, Wilson JG, Rich SS, Morrison AC, Benjamin EJ, Gross MD and Reiner AP; Cohorts for Heart and Aging Research in Genomic Epidemiology; National Heart, Lung, and Blood Institute GO Exome Sequencing Project. Association of exome sequences with plasma C-reactive protein levels in >9000 participants. Hum Mol Genet 2015; 24: 559-571.

[47] Prasad G, Giri AK, Basu A, Tandon N and Bharadwaj D. Genomewide association study for C-reactive protein in Indians replicates known associations of common variants. J Genet 2019; 98: 20.

[48] Chung RH and Chen YE. A two-stage random forest-based pathway analysis method. PLoS One 2012; 7: e36662.

[49] Montazeri M and Beigzadeh A. Machine learning models in breast cancer survival prediction. Technol Health Care 2016; 24: 31-42.

[50] Pang H, Lin A, Holford M, Enerson BE, Lu B, Lawton MP, Floyd E and Zhao H. Pathway analysis using random forests classification and regression. Bioinformatics 2006; 22: 2028-2036.

[51] Chang JS, Yeh RF, Wiencke JK, Wiemels JL, Smirnov I, Pico AR, Tihan T, Patoka J, Miike R, Sison JD, Rice T and Wrensch MR. Pathway analysis of single-nucleotide polymorphisms potentially associated with glioblastoma multiforme susceptibility using random forests. Cancer Epidemiol Biomarkers Prev 2008; 17: 1368-1373.

[52] Gawel SH, Lucht M, Gomer H, Treado P, Christensen IJ, Nielsen HJ and Davis GJ; Danish Research Group on Early Detection of Colorectal Cancer. Evaluation of algorithm development approaches: development of biomarker panels for early detection of colorectal lesions. Clin Chim Acta 2019; 498: 108-115.

[53] Mogensen UB, Ishwaran H and Gerds TA. Evaluating random forests for survival analysis using prediction error curves. J Stat Softw 2012; 50: 1-23.

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

[55] Lou XY, Chen GB, Yan L, Ma JZ, Zhu J, Elston RC and Li MD. A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. Am J Hum Genet 2007; 80: 1125-1137.

[56] Hou TT, Lin F, Bai S, Cleves MA, Xu HM and Lou XY. Generalized multifactor dimensionality reduction approaches to identification of genetic interactions underlying ordinal traits. Genet Epidemiol 2019; 43: 24-36.

[57] Xu HM, Xu LF, Hou TT, Luo LF, Chen GB, Sun XW and Lou XY. GMDR: versatile software for detecting gene-gene and gene-environment interactions underlying complex traits. Curr Genomics 2016; 17: 396-402.

[58] Xu HM, Xu LF, Hou TT, Luo LF, Chen GB, Sun XW and Lou XY. GMDR: versatile software for detecting gene-gene and gene-environment interactions underlying complex traits. 2019; http://ibi.zju.edu.cn/software/GMDR/download.html.

[59] Design of the Women's Health Initiative clinical trial and observational study. The Women''s Health Initiative Study Group. Control Clin Trials 1998; 19: 61-109.

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

[61] Golkhalkhali B, Rajandram R, Paliany AS, Ho GF, Wan Ishak WZ, Johari CS and Chin KF. Strain-specific probiotic (microbial cell preparation) and omega-3 fatty acid in modulating quality of life and inflammatory markers in colorectal cancer patients: a randomized controlled trial. Asia Pac J Clin Oncol 2018; 14: 179-191.

[62] Miranda DO, Anatriello E, Azevedo LR, Cordeiro JFC, Peria FM, Flória-Santos M and Pereira-da-Silva G. Elevated serum levels of proinflammatory cytokines potentially correlate with depression and anxiety in colorectal cancer patients in different stages of the antitumor therapy. Cytokine 2018; 104: 72-77.

[63] Nielsen FH. Dietary magnesium and chronic disease. Adv Chronic Kidney Dis 2018; 25: 230-235.

[64] Cho YA, Lee J, Oh JH, Chang HJ, Sohn DK, Shin A and Kim J. Inflammatory dietary pattern, IL-17F genetic variant, and the risk of colorectal cancer. Nutrients 2018; 10: 724.

[65] National Cancer Institute. SEER Program: Comparative Staging Guide For CancerJune 1993: https://seer.cancer.gov/archive/manuals/historic/comp_stage1.1.pdf.

[66] Jung SY, Mancuso N, Yu H, Papp J, Sobel E and 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: 31-42.

[67] 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, Nordström T, Pharoah P, Pashayan N, Schleutker J, Tammela TLJ, Sipeky C, Auvinen A, Albanes D, Weinstein S, Wolk A, Håkansson 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, Gómez-Caamaño A, Szulkin R, Eklund M, Kogevinas M, Llorca J, Castaño-Vinyals G, Penney KL, Stampfer M, Park JY, Sellers TA, Lin HY, Stanford JL, Cybulski C, Wokolorczyk D, Lubinski J, Ostrander EA, Geybels MS, Nordestgaard BG, Nielsen SF, Weischer M, Bisbjerg R, Røder MA, Iversen P, Brenner H, Cuk K, Holleczek B, Maier C, Luedeke M, Schnoeller T, Kim J, Logothetis CJ, John EM, Teixeira MR, Paulo P, Cardoso M, Neuhausen SL, Steele L, Ding YC, De Ruyck K, De Meerleer G, Ost P, Razack A, Lim J, Teo SH, Lin DW, Newcomb LF, Lessel D, Gamulin M, Kulis T, Kaneva R, Usmani N, Singhal S, Slavov C, Mitev V, Parliament M, Claessens F, Joniau S, Van den Broeck T, Larkin S, Townsend PA, Aukim-Hastie C, Gago-Dominguez M, Castelao JE, Martinez ME, Roobol MJ, Jenster G, van Schaik RHN, Menegaux F, Truong T, Koudou YA, Xu J, Khaw KT, Cannon-Albright L, Pandha H, Michael A, Thibodeau SN, McDonnell SK, Schaid DJ, Lindstrom S, Turman C, Ma J, Hunter DJ, Riboli E, Siddiq A, Canzian F, Kolonel LN, Le Marchand L, Hoover RN, Machiela MJ, Cui Z, Kraft P, Amos CI, Conti DV, Easton DF, Wiklund F, Chanock SJ, Henderson BE, Kote-Jarai Z, Haiman CA and Eeles RA; Profile Study; Australian Prostate Cancer BioResource (APCB); IMPACT Study; Canary PASS Investigators; Breast and Prostate Cancer Cohort Consortium (BPC3); PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium; Cancer of the Prostate in Sweden (CAPS); Prostate Cancer Genome-wide Association Study of Uncommon Susceptibility Loci (PEGASUS); Genetic Associations and Mechanisms in Oncology (GAME-ON)/Elucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE) Consortium. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet 2018; 50: 928-936.

[68] Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M and Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics 2010; 26: 2867-2873.

[69] Ishwaran H and Kogalur UB. Random Survival Forests for R. 2007. https://pdfs.semanticscholar.org/951a/84f0176076fb6786fdf43320e8b27094dcfa.pdf.

[70] Harrell FE Jr, Califf RM, Pryor DB, Lee KL and Rosati RA. Evaluating the yield of medical tests. JAMA 1982; 247: 2543-2546.

[71] Wiens BL, Dmitrienko A and Marchenko O. Selection of hypothesis weights and ordering when testing multiple hypotheses in clinical trials. J Biopharm Stat 2013; 23: 1403-1419.

[72] Gene Card: Human Gene Database: One Cut Homeobox 2. 2020; https://www.genecards.org/cgi-bin/carddisp.pl?gene=ONECUT2.

[73] Lu T, Wu B, Yu Y, Zhu W, Zhang S, Zhang Y, Guo J and Deng N. Blockade of ONECUT2 expression in ovarian cancer inhibited tumor cell proliferation, migration, invasion and angiogenesis. Cancer Sci 2018; 109: 2221-2234.

[74] Zhang J, Cheng J, Zeng Z, Wang Y, Li X, Xie Q, Guo J and Deng N. Comprehensive profiling of novel microRNA-9 targets and a tumor suppressor role of microRNA-9 via targeting IGF2BP1 in hepatocellular carcinoma. Oncotarget 2015; 6: 42040-42052.

[75] Zhang K, Wang YW, Wang YY, Song Y, Zhu J, Si PC and Ma R. Identification of microRNA biomarkers in the blood of breast cancer patients based on microRNA profiling. Gene 2017; 619: 10-20.

[76] Guo H, Ci X, Ahmed M, Hua JT, Soares F, Lin D, Puca L, Vosoughi A, Xue H, Li E, Su P, Chen S, Nguyen T, Liang Y, Zhang Y, Xu X, Xu J, Sheahan AV, Ba-Alawi W, Zhang S, Mahamud O, Vellanki RN, Gleave M, Bristow RG, Haibe-Kains B, Poirier JT, Rudin CM, Tsao MS, Wouters BG,

Fazli L, Feng FY, Ellis L, van der Kwast T, Berlin A, Koritzinsky M, Boutros PC, Zoubeidi A, Beltran H, Wang Y and He HH. ONECUT2 is a driver of neuroendocrine prostate cancer. Nat Commun 2019; 10: 278.

[77] Sun Y, Shen S, Liu X, Tang H, Wang Z, Yu Z, Li X and Wu M. MiR-429 inhibits cells growth and invasion and regulates EMT-related marker genes by targeting Onecut2 in colorectal carcinoma. Mol Cell Biochem 2014; 390: 19-30.

[78] HNF4A hepatocyte nuclear factor 4 alpha. Genetics Home Reference 2020; https://www.ncbi.nlm.nih.gov/gene/3172.

[79] Darsigny M, Babeu JP, Seidman EG, Gendron FP, Levy E, Carrier J, Darsigny M, Babeu JP, Seidman EG, Gendron FP, Levy E, Carrier J, Perreault N and Boudreau F. Hepatocyte nuclear factor-4alpha promotes gut neoplasia in mice and protects against the production of reactive oxygen species. Cancer Res 2010; 70: 9423-9433.

[80] López-Mejías R, Genre F, Remuzgo-Martínez S, González-Juanatey C, Robustillo-Villarino M, Llorca J, Corrales A, Vicente E, Miranda-Filloy JA, Magro C, Tejera-Segura B, Ramírez Huaranga MA, Pina T, Blanco R, Alegre-Sancho JJ, Raya E, Mijares V, Ubilla B, Mínguez Sánchez MD, Gómez-Vaquero C, Balsa A, Pascual-Salcedo D, López-Longo FJ, Carreira P, González-Álvaro I, Rodríguez-Rodríguez L, Fernández-Gutiérrez B, Ferraz-Amaro I, Castañeda S, Martín J and González-Gay MA. Influence of elevated-CRP level-related polymorphisms in non-rheumatic Caucasians on the risk of subclinical atherosclerosis and cardiovascular disease in rheumatoid arthritis. Sci Rep 2016; 6: 31979.

[81] Ligthart S, de Vries PS, Uitterlinden AG, Hofman A; CHARGE Inflammation working group, Franco OH, Chasman DI and Dehghan A. Pleiotropy among common genetic loci identified for cardiometabolic disorders and C-reactive protein. PLoS One 2015; 10: e0118859.

[82] Gene Card: human gene database: hepatocyte nuclear factor 4 alpha. 2020; https://www.genecards.org/cgi-bin/carddisp.pl?gene=HNF4A.

[83] Babeu JP and Boudreau F. Hepatocyte nuclear factor 4-alpha involvement in liver and intestinal inflammatory networks. World J Gastroenterol 2014; 20: 22-30.

[84] Yang YC, Fu WP, Zhang J, Zhong L, Cai SX and Sun C. rs401681 and rs402710 confer lung cancer susceptibility by regulating TERT expression instead of CLPTM1L in East Asian populations. Carcinogenesis 2018; 39: 1216-1221.

[85] Selvaraj G, Kaliamurthi S, Kaushik AC, Khan A, Wei YK, Cho WC, Gu K and Wei DQ. Identification of target gene and prognostic evaluation for lung adenocarcinoma using gene expression meta-analysis, network analysis and neural network algorithms. J Biomed Inform 2018; 86: 120-134.

[86] Tentu S, Nandarapu K, Muthuraj P, Venkitasamy K, Venkatraman G and Rayala SK. DHQZ-17, a potent inhibitor of the transcription factor HNF4A, suppresses tumorigenicity of head and neck squamous cell carcinoma in vivo. J Cell Physiol 2018; 233: 2613-2628.

[87] Rogerson C, Britton E, Withey S, Hanley N, Ang YS and Sharrocks AD. Identification of a primitive intestinal transcription factor network shared between esophageal adenocarcinoma and its precancerous precursor state. Genome Res 2019; 29: 723-736.

[88] Li MM, Tang YQ, Gong YF, Cheng W, Li HL, Kong FE, Guan XY, Ma NF and Liu M. Development of an oncogenic dedifferentiation SOX signature with prognostic significance in hepatocellular carcinoma. BMC Cancer 2019; 19: 851.

[89] Takashima Y, Horisawa K, Udono M, Ohkawa Y and Suzuki A. Prolonged inhibition of hepatocellular carcinoma cell proliferation by combinatorial expression of defined transcription factors. Cancer Sci 2018; 109: 3543-3553.

[90] Yao HS, Wang J, Zhang XP, Wang LZ, Wang Y, Li XX, Jin KZ, Hu ZQ and Wang WJ. Hepatocyte nuclear factor 4alpha suppresses the aggravation of colon carcinoma. Mol Carcinog 2016; 55: 458-472.

[91] Shabana, Shahid SU and Hasnain S. Use of a gene score of multiple low-modest effect size variants can predict the risk of obesity better than the individual SNPs. Lipids Health Dis 2018; 17: 155.

[92] Salva-Pastor N, Chavez-Tapia NC, Uribe M and Nuno-Lambarri N. Understanding the association of polycystic ovary syndrome and non-alcoholic fatty liver disease. J Steroid Biochem Mol Biol 2019; 194: 105445.

[93] Simo R, Barbosa-Desongles A, Lecube A, Hernandez C and Selva DM. Potential role of tumor necrosis factor-alpha in downregulating sex hormone-binding globulin. Diabetes 2012; 61: 372-382.

[94] Slattery ML, Potter JD, Curtin K, Edwards S, Ma KN, Anderson K, Schaffer D and Samowitz WS. Estrogens reduce and withdrawal of estrogens increase risk of microsatellite instability-positive colon cancer. Cancer Res 2001; 61: 126-130.

[95] Issa JP. Colon cancer: it's CIN or CIMP. Clin Cancer Res 2008; 14: 5939-5940.

[96] Martínez ME, Grodstein F, Giovannucci E, Colditz GA, Speizer FE, Hennekens C, Rosner B, Willett WC and Stampfer MJ. A prospective study of reproductive factors, oral contraceptive use,

and risk of colorectal cancer. Cancer Epidemiol 1997; 6: 1-5.

[97] Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, Macera CA, Heath GW, Thompson PD and Bauman A. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007; 39: 1423-1434.

[98] Van Horn L, Carson JA, Appel LJ, Burke LE, Economos C, Karmally W, Lancaster K, Lichtenstein AH, Johnson RK, Thomas RJ, Vos M, Wylie-Rosett J and Kris-Etherton P; American Heart Association Nutrition Committee of the Council on Lifestyle and Cardiometabolic Health; Council on Cardiovascular Disease in the Young; Council on Cardiovascular and Stroke Nursing; Council on Clinical Cardiology; and Stroke Council. Recommended dietary pattern to achieve adherence to the American Heart Association/American College of Cardiology (AHA/ACC) guidelines: a scientific statement from the American Heart Association. Circulation 2016; 134: e505-e529.

[99] Waist Circumference and Waist-Hip Ratio. The World Health Organization; 2011. https://apps.who.int/iris/bitstream/handle/10665/44583/9789241501491_eng.pdf;jsessionid=587F1A4B31954337F34F1A84F063AB0E?sequence=1.

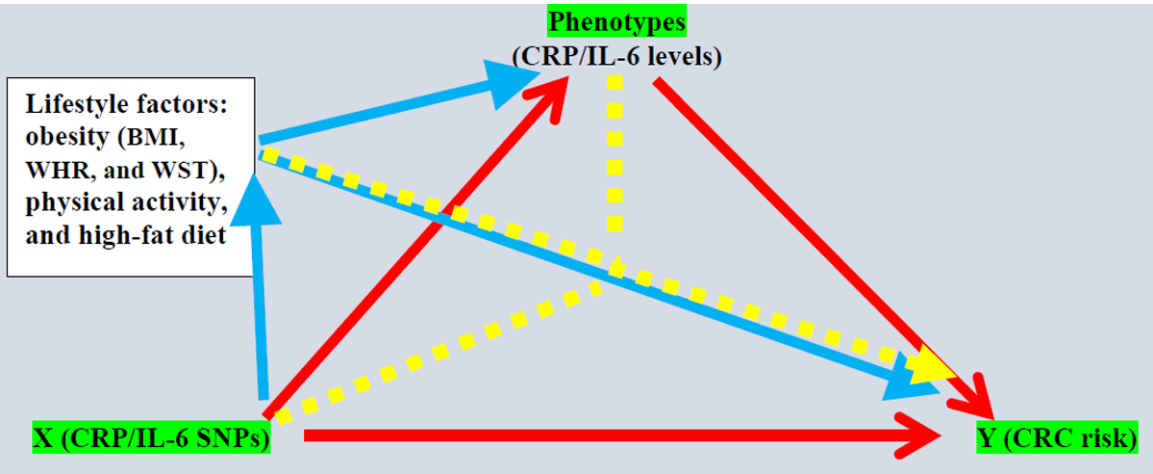


Figure S1. Empirical pathways of pro-inflammatory SNPs, phenotypes, and CRC risk, interplaying with obesity status and obesity-lifestyle factors (Note: pathways in red and blue lines were tested in our previous GWA G×E and current post-GWA analyses; yellow lines reflect conceptual framework. BMI, body mass index; CRC, colorectal cancer; CRP, C-reactive protein; G×E, gene-environment interaction; GWA, genome-wide analysis; IL-6, interleukin-6; SNP, single-nucleotide polymorphism; WHR, waist-to-hip ratio; WST, waist circumference).

Table S1A. Allele frequencies (total n=10,798) of the 87 SNPs associated with pro-inflammatory phenotypes from our previous GWA analysis

Chr	Position¥	Gene	SNP	Allele		Alt Allele Frequency					
						AS264	GARNET	GECCOCYTO	GECCOINIT	HIPFX	WHIMS
				Ref	Alt	n=1,603	n=2,382	n=1,177	n=216	n=1,909	n=3,511
1	159652939	CRPP1	rs2592887	C	T	0.41	0.41	0.41	0.38	0.40	0.39
1	159653599	CRPP1	rs1470515	C	T	0.40	0.39	0.40	0.37	0.38	0.38
1	159655726	CRPP1	rs2592902	G	T	0.40	0.39	0.40	0.37	0.38	0.38
1	159665921	CRPP1	rs2808624	C	G	0.40	0.39	0.40	0.36	0.38	0.38
1	159668984	CRPP1	rs11265257	C	T	0.40	0.39	0.40	0.36	0.38	0.38
1	159674933	CRPP1	rs876537	C	T	0.40	0.39	0.40	0.36	0.38	0.38
1	159676011	CRPP1	rs2808628	G	A	0.34	0.34	0.34	0.32	0.33	0.33
1	159676796	CRPP1	rs2808629	G	A	0.34	0.34	0.35	0.32	0.33	0.33
1	159678816	CRPP1/CRP	rs2794520	C	T	0.34	0.34	0.34	0.32	0.33	0.33
1	159682233	CRP	rs1205	C	T	0.34	0.34	0.34	0.32	0.33	0.33
1	159684665	CRP	rs3091244	G	A	0.32	0.36	0.32	0.34	0.34	0.37
1	159689388	CRP	rs2027471	T	A	0.35	0.35	0.35	0.33	0.33	0.33
1	159691559	CRP	rs1341665	G	A	0.35	0.35	0.35	0.33	0.33	0.33
1	159693605	CRP	rs2211320	G	A	0.33	0.33	0.33	0.31	0.31	0.32
1	159694779	CRP	rs7551731	T	C	0.34	0.34	0.34	0.32	0.32	0.33
1	159698549	CRP	rs7553007	G	A	0.34	0.34	0.34	0.31	0.32	0.32
1	159699249	CRP	rs4546916	G	T	0.34	0.34	0.34	0.31	0.32	0.32
1	159703442	CRP	rs4287174	T	A	0.34	0.34	0.34	0.31	0.32	0.32
1	159703462	CRP	rs4428887	A	G	0.35	0.34	0.35	0.32	0.33	0.33
1	159706230	CRP	rs12037186	A	G	0.33	0.33	0.33	0.30	0.31	0.32
1	159708825	CRP/RP11-419N10.5	rs12042360	G	A	0.16	0.17	0.18	0.16	0.16	0.15
1	159713844	CRP/RP11-419N10.5	rs12049404	C	T	0.16	0.17	0.18	0.16	0.16	0.15
1	159717162	CRP/RP11-419N10.5	rs11588887	G	A	0.16	0.17	0.18	0.16	0.16	0.15
9	118330052	DEC1	rs149109490	T	C	0.99	0.99	0.99	0.99	0.99	0.99
12	121380544	HNF1A-AS1	rs2649999	T	C	NA	0.66	0.67	NA	0.68	0.65
12	121384495	HNF1A-AS1	rs11065358	T	C	NA	0.63	0.63	NA	0.64	0.62
12	121388559	HNF1A-AS1	rs1696359	T	C	NA	0.65	0.66	NA	0.67	0.64
12	121388962	HNF1A-AS1	rs2650000	A	C	0.66	0.65	0.66	NA	0.67	0.65
12	121390078	HNF1A-AS1	rs2701194	A	G	NA	0.63	0.65	NA	0.66	0.63
12	121391671	HNF1A-AS1	rs2701175	C	A	0.63	0.61	0.63	NA	0.64	0.61
12	121392040	HNF1A-AS1	rs11065365	G	A	NA	0.55	0.56	NA	0.57	0.54
12	121392341	HNF1A-AS1	rs1732391	C	T	0.66	0.65	0.66	NA	0.67	0.65

12	121397875	HNF1A-AS1	rs6489786	A	G	0.66	0.65	0.66	NA	0.67	0.65
12	121398654	HNF1A-AS1	rs7954039	A	C	0.66	0.65	0.66	NA	0.67	0.65
12	121398657	HNF1A-AS1	rs7954331	G	T	0.66	0.65	0.66	NA	0.67	0.65
12	121403724	HNF1A-AS1	rs7953249	G	A	0.59	0.58	0.59	NA	0.60	0.58
12	121404155	HNF1A-AS1	rs7135337	A	C	0.58	0.57	0.58	NA	0.58	0.57
12	121404584	HNF1A-AS1	rs1920792	C	T	0.51	0.53	0.53	NA	0.51	0.53
12	121405126	HNF1A-AS1	rs2251468	C	A	0.66	0.65	0.66	NA	0.67	0.65
12	121405210	HNF1A-AS1	rs10774579	C	T	0.51	0.53	0.52	NA	0.51	0.53
12	121406293	HNF1A-AS1	rs2393792	A	G	0.50	0.53	0.52	NA	0.51	0.53
12	121406370	HNF1A-AS1	rs2243616	G	T	0.65	0.64	0.65	NA	0.66	0.63
12	121413027	HNF1A-AS1	rs148608463	A	G	0.66	0.65	0.66	NA	0.67	0.65
12	121413345	HNF1A-AS1	rs142632970	G	A	0.68	0.67	0.68	NA	0.69	0.67
12	121414915	HNF1A-AS1	rs2255531	A	G	0.65	0.64	0.65	NA	0.67	0.64
12	121415293	HNF1A-AS1	rs7139079	G	A	0.59	0.58	0.60	NA	0.60	0.58
12	121415390	HNF1A-AS1	rs2464190	C	T	0.58	0.55	0.59	NA	0.58	0.56
12	121416622	HNF1A	rs1169289	G	C	0.57	0.54	0.58	NA	0.57	0.55
12	121416650	HNF1A	rs1169288	C	A	0.68	0.67	0.68	NA	0.69	0.67
12	121416988	HNF1A	rs2244608	G	A	0.68	0.67	0.68	NA	0.70	0.67
12	121419056	HNF1A	rs1169286	C	T	0.58	0.56	0.57	NA	0.59	0.57
12	121419926	HNF1A	rs1169284	C	T	0.68	0.68	0.69	NA	0.70	0.68
12	121420260	HNF1A	rs7979473	A	G	0.60	0.61	0.62	NA	0.61	0.61
12	121420263	HNF1A	rs7979478	A	G	0.60	0.60	0.61	NA	0.61	0.61
12	121420807	HNF1A	rs1183910	A	G	0.69	0.68	0.69	NA	0.70	0.68
12	121423285	HNF1A	rs11065384	T	C	0.68	0.68	0.69	NA	0.70	0.68
12	121423376	HNF1A	rs7970695	G	A	0.62	0.61	0.62	NA	0.63	0.61
12	121423386	HNF1A	rs11065385	A	G	0.69	0.68	0.69	NA	0.70	0.68
12	121423659	HNF1A	rs9738226	A	G	0.62	0.61	0.62	NA	0.63	0.61
12	121423956	HNF1A	rs2393791	C	T	0.62	0.61	0.62	NA	0.63	0.61
12	121424406	HNF1A	rs2393776	G	A	0.62	0.61	0.62	0.60	0.63	0.61
12	121424490	HNF1A	rs2243458	T	C	0.69	0.68	0.69	0.66	0.70	0.68
12	121424574	HNF1A	rs2393775	G	A	0.62	0.61	0.62	0.60	0.63	0.61
12	121424861	HNF1A	rs7310409	A	G	0.62	0.61	0.62	0.59	0.63	0.61
12	121426478	HNF1A	rs1169292	T	C	0.69	0.68	0.69	0.67	0.70	0.68
12	121426594	HNF1A	rs1169294	A	G	0.69	0.68	0.69	0.67	0.70	0.68
12	121431225	HNF1A	rs1169300	A	G	0.70	0.69	0.71	0.69	0.72	0.70
12	121431300	HNF1A	rs1169301	T	C	0.70	0.69	0.71	0.69	0.72	0.70

12	121432603	HNF1A	rs2264782	T	C	0.65	0.63	0.66	0.63	0.66	0.65
12	121434833	HNF1A	rs2259852	A	G	0.65	0.63	0.66	0.63	0.66	0.64
12	121435342	HNF1A	rs2259820	T	C	0.70	0.69	0.71	0.69	0.72	0.70
12	121435427	HNF1A	rs2464196	A	G	0.70	0.69	0.71	0.69	0.72	0.70
12	121435475	HNF1A	rs2464195	A	G	0.65	0.63	0.66	0.63	0.66	0.64
12	121435587	HNF1A	rs2259816	T	G	0.65	0.63	0.66	0.66	0.66	0.64
12	121438311	HNF1A	rs1169306	T	C	0.65	0.63	0.66	0.63	0.66	0.64
12	121438844	HNF1A	rs735396	C	T	0.65	0.63	0.66	0.63	0.66	0.64
12	121439192	HNF1A	rs1169309	T	G	0.65	0.63	0.66	0.66	0.66	0.64
12	121439433	HNF1A	rs1169310	A	G	0.65	0.63	0.66	0.63	0.66	0.64
12	121440731	C12orf43	rs1169311	T	C	0.65	0.63	0.66	0.63	0.66	0.64
12	121441461	C12orf43	rs1169312	T	G	0.65	0.63	0.66	0.63	0.66	0.64
12	121442670	C12orf43	rs1169313	C	T	0.65	0.63	0.65	0.63	0.66	0.64
12	121445808	C12orf43	rs2257962	C	T	0.65	0.63	0.65	0.63	0.66	0.64
12	121450384	C12orf43	rs2254971	C	C	0.63	0.61	0.64	0.60	0.63	0.62
12	121454622	C12orf43	rs1182933	T	C	0.70	0.69	0.71	0.69	0.72	0.70
19	45411941	APOE	rs429358	C	T	0.85	0.87	0.87	NA	0.87	0.86
22	22190785	MAPK1	rs56398890	A	T	0.58	0.56	0.56	0.53	0.57	0.56
22	22202164	MAPK1	rs9607320	T	C	0.59	0.58	0.57	0.53	0.58	0.57

Alt, alternative; Chr, chromosome; GWA, genome-wide association; Ref, reference; SNP, single-nucleotide polymorphism. ¥GRCh 37 coordinated.

Table S1B. Allele frequencies (total n=10,798) of the 65 SNPs associated with pro-inflammatory phenotypes from other GWA studies

Chr	Positions¥	Gene	SNP	Allele		Alt Allele Frequency					
						AS264	Garnet	Geccocyto	Geccoinit	Hipfx	Whims
				Ref	Alt	n=1,603	n=2,382	n=1,177	n=216	n=1,909	n=3,511
1	27180088	ZDHHC18	rs75460349	A	C	0.02	0.03	0.02	0.02	0.03	0.02
1	40036847	PABPC4	rs2293476	G	C	0.21	0.22	0.24	0.25	0.23	0.23
1	40064961	PABPC4/HEYL	rs12037222	G	A	NA	0.22	0.24	0.24	0.23	0.23
1	66085574	LEPR	rs3790439	A	T	0.63	0.63	0.62	0.67	0.63	0.63
1	66102257	LEPR	rs1805096	A	G	0.63	0.62	0.62	0.67	0.63	0.63
1	66161461	LEPR	rs4420065	T	C	0.63	0.62	0.62	0.68	0.63	0.63
1	91530305	ZNF644	rs469772	C	T	0.20	0.20	0.19	0.19	0.18	0.19
1	154426264	IL6R	rs4129267	C	T	0.41	0.40	0.41	0.40	0.40	0.41
1	154426970	IL6R	rs2228145	A	C	0.41	0.40	0.41	0.40	0.41	0.41
1	159683438	CRP	rs1800947	C	G	0.06	0.06	NA	0.06	0.06	0.05

1	159684186	CRP	rs1417938	T	A	0.30	0.31	0.31	0.32	0.32	0.32
1	247601595	NLRP3	rs12239046	C	T	0.37	0.37	0.38	0.36	0.37	0.37
1	247612562	NLRP3	rs10925027	C	T	0.40	0.40	NA	0.42	0.40	0.39
2	629881	TMEM18	rs12995480	C	T	0.17	0.18	0.18	0.17	0.17	0.18
2	27730940	GCKR	rs1260326	C	T	0.41	0.41	0.41	NA	0.41	0.40
2	88438050	FABP1	rs4246598	C	A	0.45	0.46	0.45	0.47	0.45	0.46
2	102744854	IL1R1	rs9284725	A	C	0.24	0.25	0.23	0.25	0.25	0.25
2	113838145	IL1F10	rs13409371	G	A	0.39	0.39	0.41	0.43	0.39	0.40
2	113841030	IL1F10	rs6734238	A	G	0.40	0.40	0.42	0.44	0.40	0.40
2	214033530	IKZF2	rs1441169	G	A	0.49	0.49	0.48	0.50	0.48	0.48
3	49891885	TRAIP	rs2352975	C	T	NA	0.69	NA	0.70	0.69	0.69
3	170705693	EIF5A2	rs1514895	A	G	0.30	0.30	0.30	0.27	0.28	0.29
5	131839618	IRF1	rs4705952	A	G	0.25	NA	0.27	0.24	0.25	0.24
5	172191052	DUSP1	rs17658229	T	C	0.04	NA	0.04	0.03	0.04	0.04
6	116314634	FRK	rs12202641	C	T	0.41	NA	0.41	0.41	0.40	0.41
6	117114025	GPRC6A	rs6901250	G	A	0.33	0.32	0.32	0.32	0.33	0.31
6	126851160	CENPW	rs1490384	T	C	0.49	0.51	0.51	0.54	0.47	0.49
6	130371227	L3MBTL3	rs9385532	C	T	0.33	0.31	0.31	0.33	0.32	0.33
7	22759469	IL6	rs1880241	G	A	0.52	0.52	0.50	0.52	0.51	0.51
7	22766645	IL6	rs1800795	G	C	0.43	0.43	0.41	0.40	0.43	0.44
7	36084529	EEPD1	rs2710804	T	C	0.37	0.36	0.39	0.38	0.38	0.38
7	72971231	BCL7B	rs13233571	C	T	0.11	0.13	0.11	0.13	0.12	0.12
8	9183358	PPP1R3B	rs9987289	G	A	0.09	0.09	0.08	0.08	0.09	0.09
8	9183596	PPP1R3B	rs4841132	G	A	0.09	0.09	0.08	0.08	0.09	0.09
8	117007850	TRPS1	rs2064009	T	C	0.42	0.43	0.40	0.39	0.41	0.41
8	126344208	NSMCE2	rs2891677	C	T	0.55	0.54	0.54	0.57	0.55	0.55
9	136142355	ABO	rs643434	G	A	0.36	0.37	0.35	0.31	0.35	0.35
10	91007360	LIPA	rs1051338	T	G	0.30	0.30	0.29	0.28	0.30	0.29
11	13357183	ARNTL	rs10832027	A	G	0.32	0.32	0.31	0.36	0.32	0.32
11	47312892	MADD	rs10838687	T	G	0.21	0.21	0.20	0.23	0.21	0.20
11	60021948	MS4A4A	rs1582763	G	A	0.36	0.38	0.35	0.39	0.36	0.38
11	72496148	STARD10	rs7121935	G	A	0.34	0.37	0.38	NA	0.38	0.38
12	95855385	METAP2	rs11108056	C	G	0.44	0.43	NA	0.44	0.46	0.43
12	103483094	ASCL1	rs10745954	G	A	0.52	0.52	0.50	0.55	0.51	0.52
12	103537266	C12orf42	rs10778215	A	T	0.53	0.53	0.53	0.58	0.52	0.53

14	73011885	RGS6	rs2239222	A	G	0.36	0.36	0.34	NA	0.36	0.35
14	94838142	SERPINA1/SERPINA2P	rs112635299	G	T	0.02	0.02	0.02	0.03	0.02	0.02
15	51745277	DMXL2	rs4774590	G	A	0.38	0.37	0.39	0.35	0.38	0.39
15	53728154	WDR72	rs1189402	A	G	0.37	0.37	0.39	0.38	0.36	0.38
15	60878030	RORA	rs340005	A	G	0.38	0.37	0.37	0.38	0.39	0.38
15	60894965	RORA	rs340029	T	C	0.39	0.37	0.37	0.38	0.39	0.37
16	53803574	FTO	rs1558902	T	A	0.39	0.41	0.40	0.42	0.40	0.40
17	16097430	NCOR1	rs178810	C	T	NA	0.56	0.57	0.57	0.57	0.56
17	72699833	CD300LF/RAB37	rs10512597	C	T	0.19	0.19	0.18	0.16	0.18	0.19
18	12821593	PTPN2	rs2847281	A	G	0.39	0.39	0.38	0.39	0.40	0.40
18	12841176	PTPN2	rs2852151	G	A	0.39	0.39	0.38	0.39	0.40	0.40
18	55080437	ONECUT2	rs4092465	A	G	0.62	0.63	NA	NA	0.65	0.64
18	57897803	MC4R	rs12960928	T	C	0.26	0.26	0.26	0.26	0.27	0.27
19	45395714	TOMM40	rs157581	T	C	NA	0.21	0.21	NA	0.20	0.21
19	45396144	TOMM40	rs11556505	C	T	0.13	0.13	0.12	NA	0.12	0.14
19	45396219	TOMM40	rs157582	C	T	0.23	0.21	0.20	NA	0.20	0.21
20	43042364	HNF4A	rs1800961	C	T	0.03	0.03	NA	0.04	0.03	0.03
20	62343956	ZGPAT	rs2315008	G	T	0.31	0.32	0.32	0.30	0.31	0.33
21	40465534	PSMG1	rs2836878	G	A	0.27	0.27	0.25	0.28	0.27	0.27
22	39074737	TOMM22	rs6001193	A	G	NA	0.35	0.37	0.30	0.36	0.36

Alt, alternative; Chr, chromosome; GWA, genome-wide association; Ref, reference; SNP, single-nucleotide polymorphism. ¥GRCh 37 coordinated.

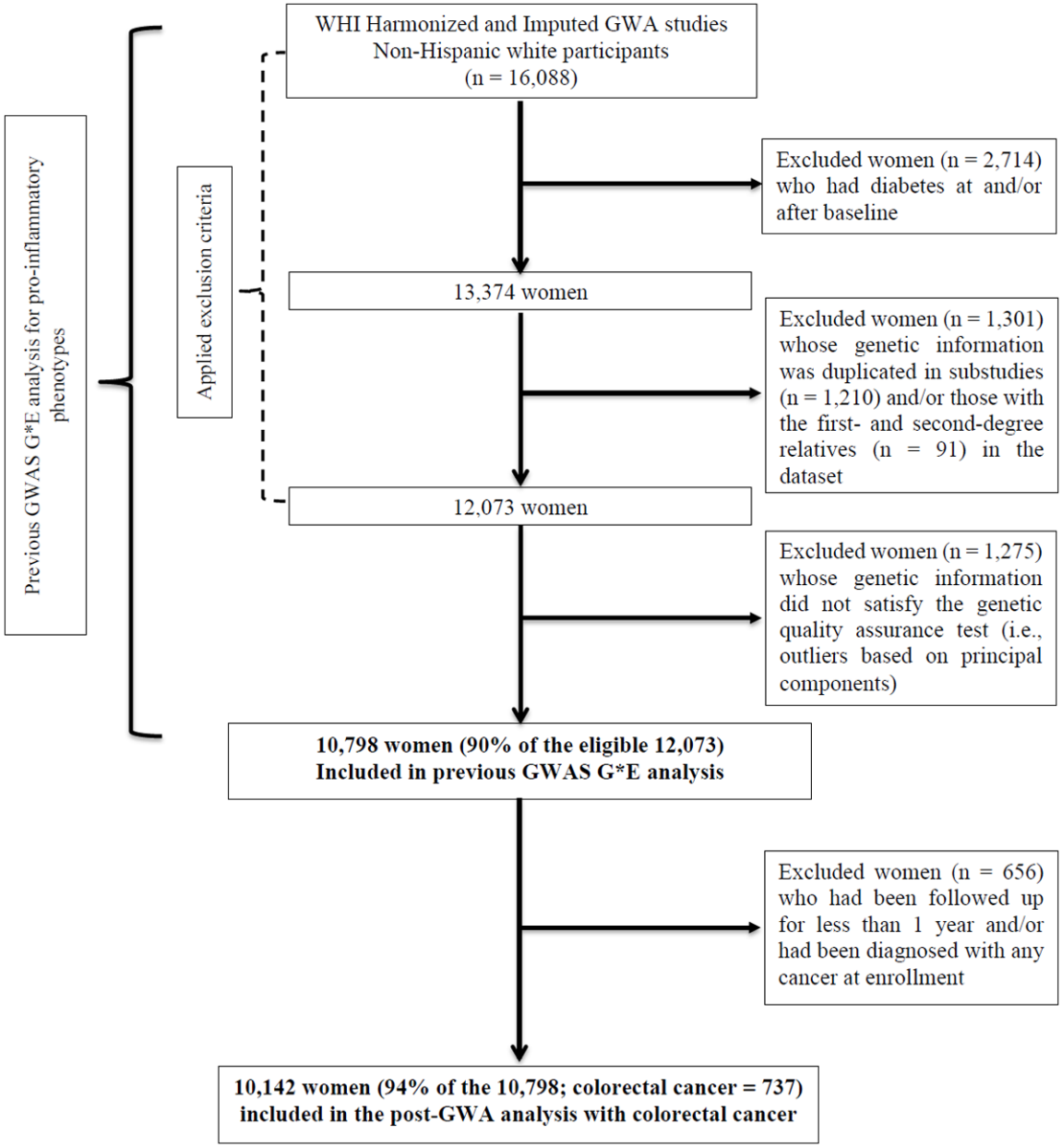
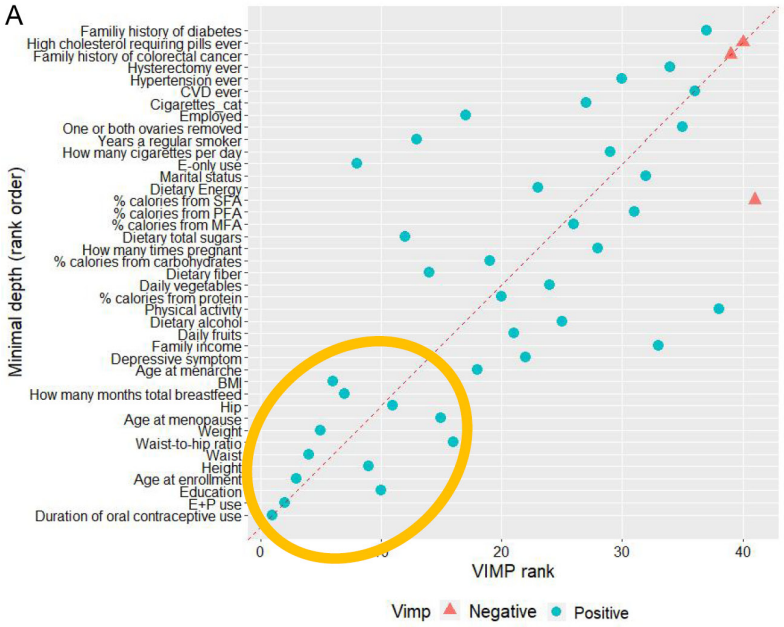


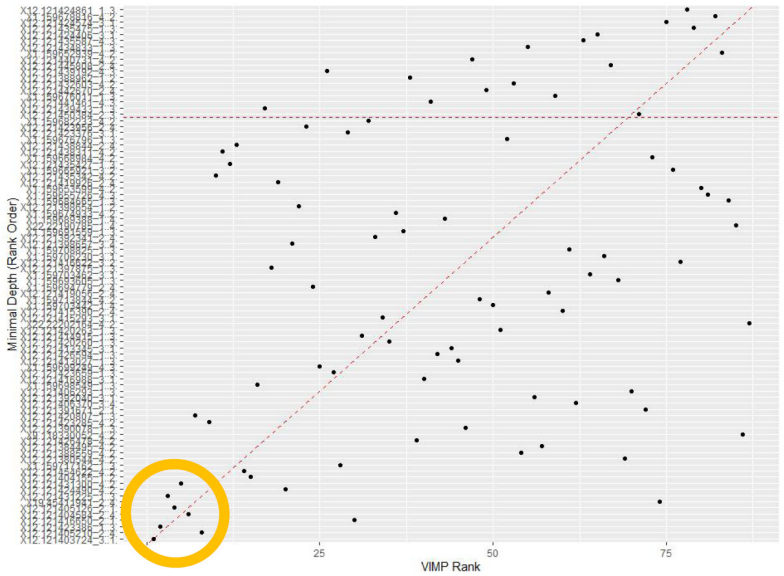
Figure S2. Flow diagram of analytic cohort derived from the previous GWA G×E interaction study. (G×E, gene-environment; GWAS, genome-wide association study; WHI, Women's Health Initiative).

A

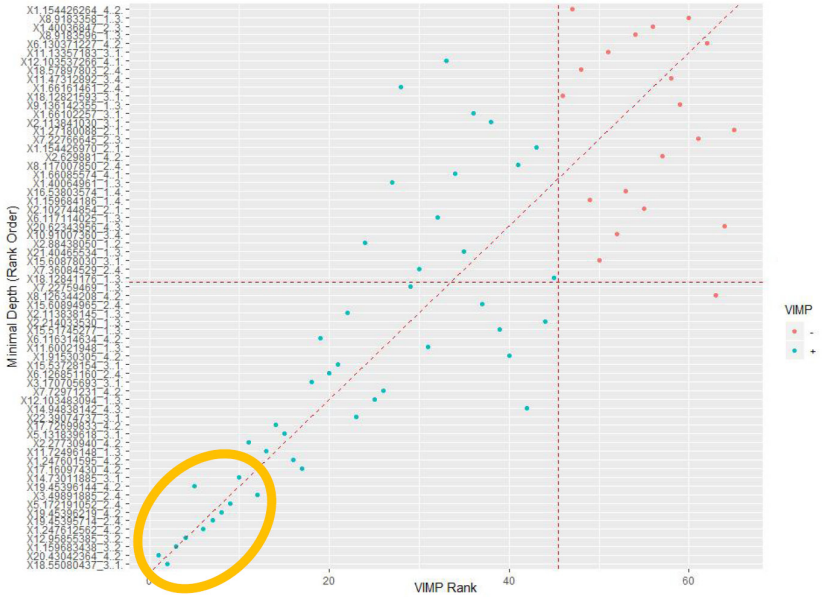


B

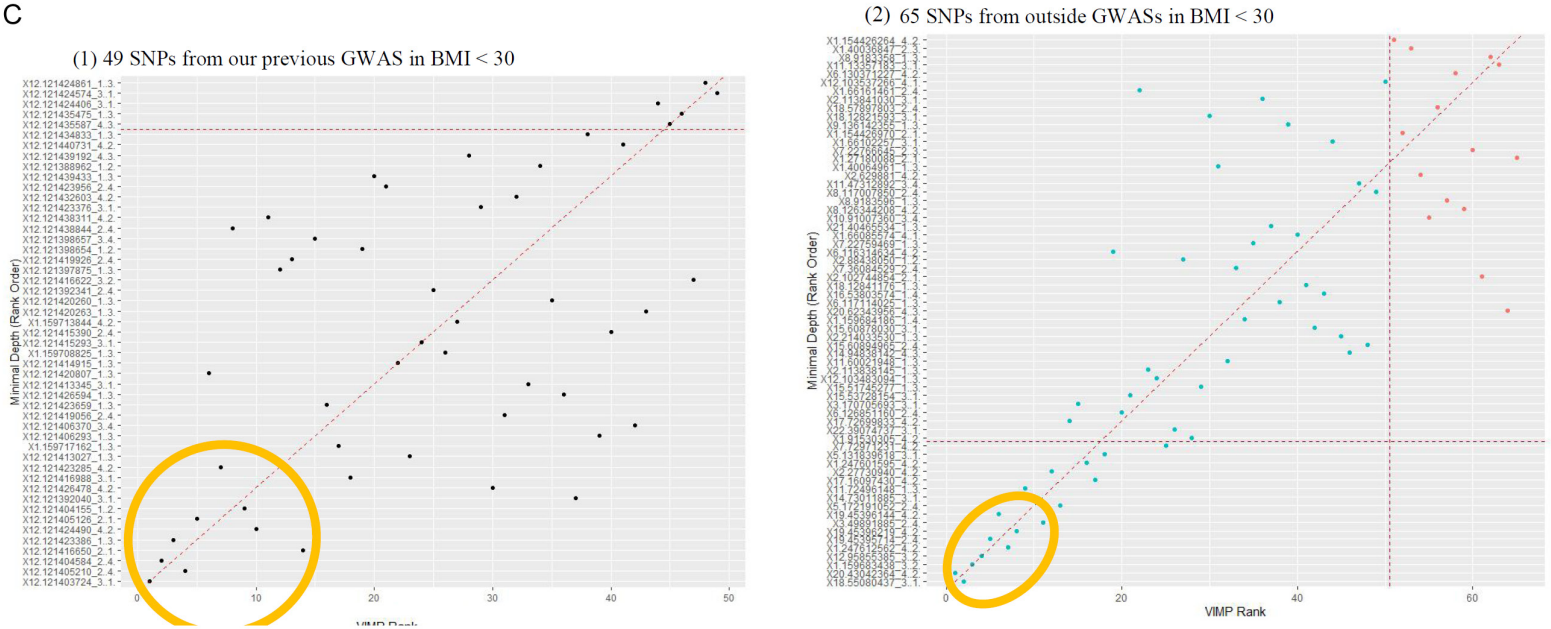
(1) 87 SNPs from our previous GWAS in overall analysis



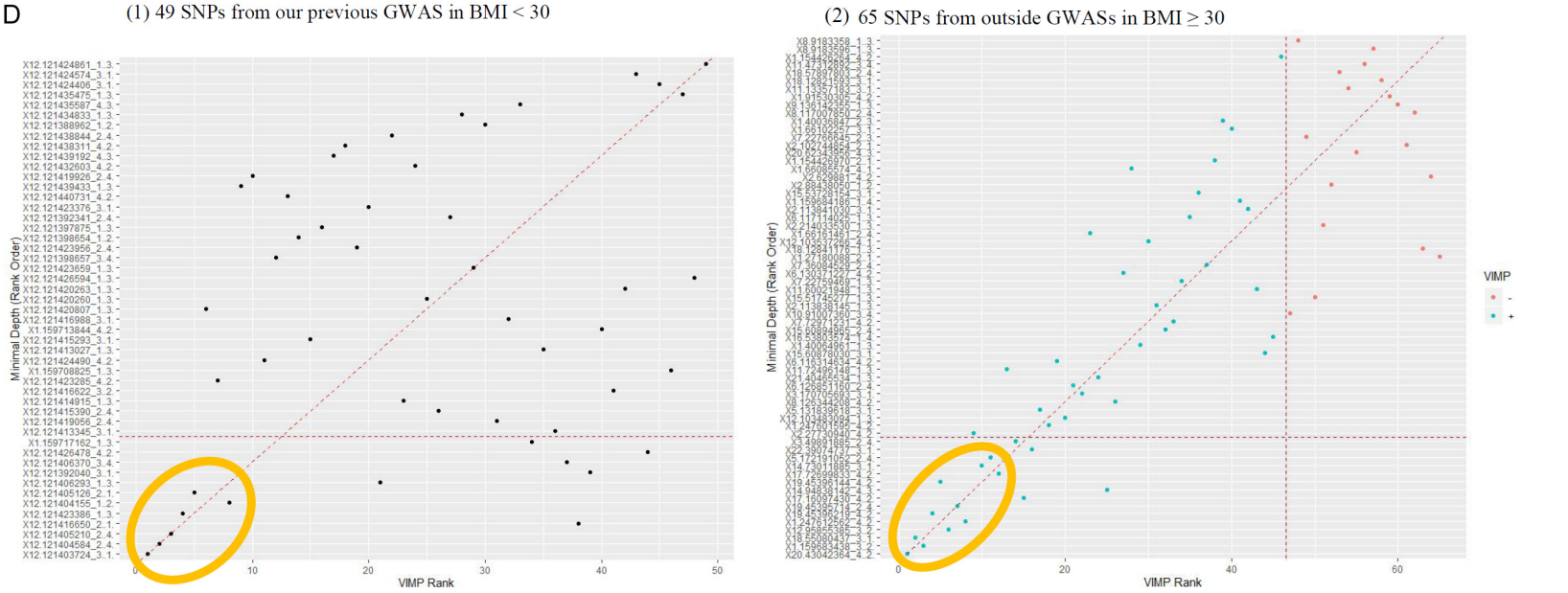
(2) 65 SNPs from outside GWASs in overall analysis



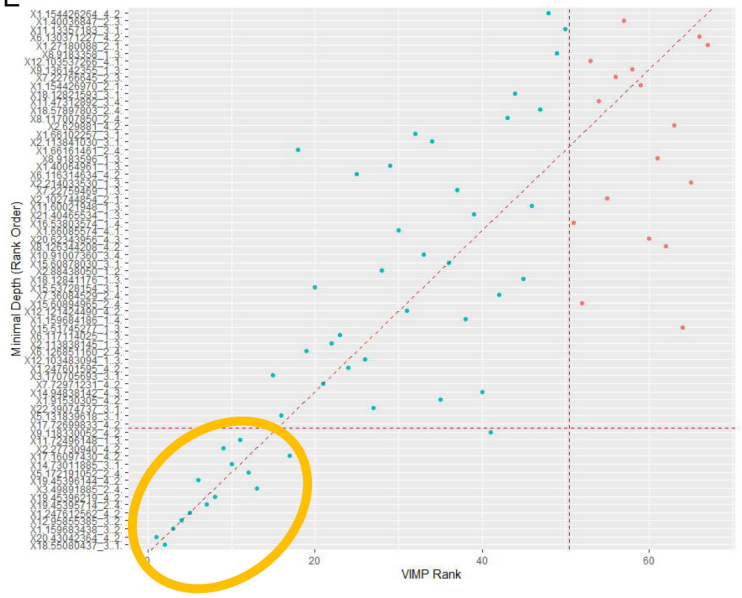
C


(1) 49 SNPs from our previous GWAS in BMI < 30
(2) 65 SNPs from outside GWASs in BMI < 30
Minimal Depth (Rank Order)
VIMP Rank
VIMP
-
+

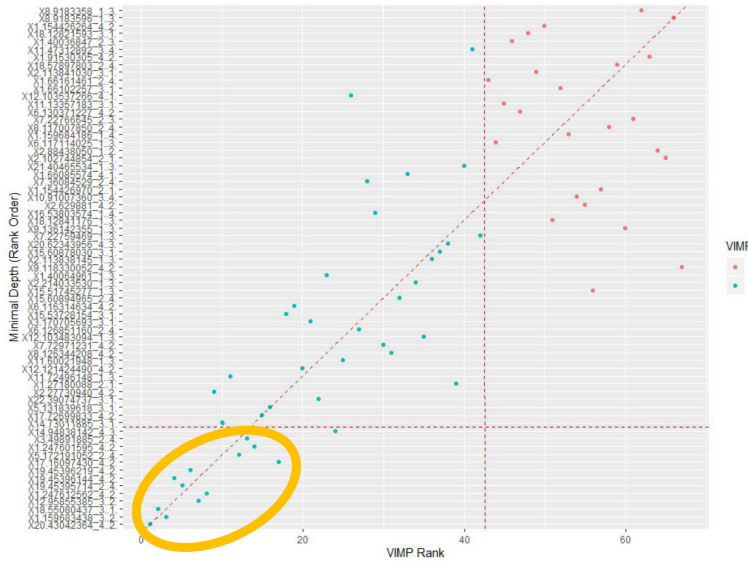
D


(1) 49 SNPs from our previous GWAS in BMI < 30
(2) 65 SNPs from outside GWASs in BMI ≥ 30
Minimal Depth (Rank Order)
VIMP Rank
VIMP
-
+

E

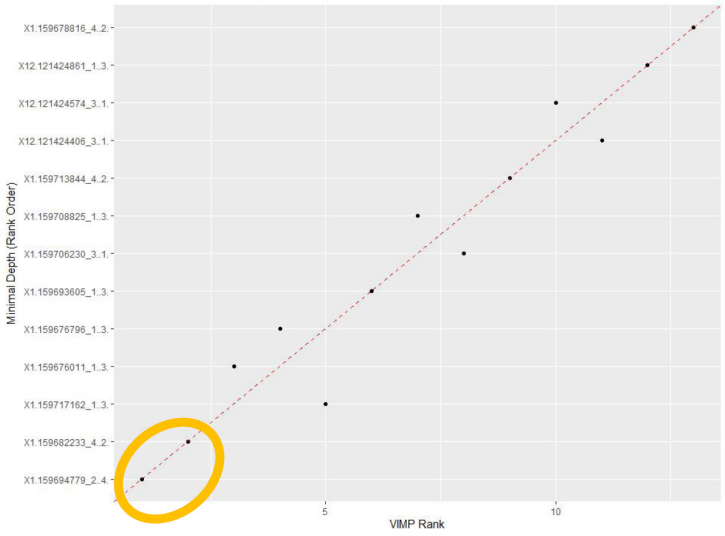


F

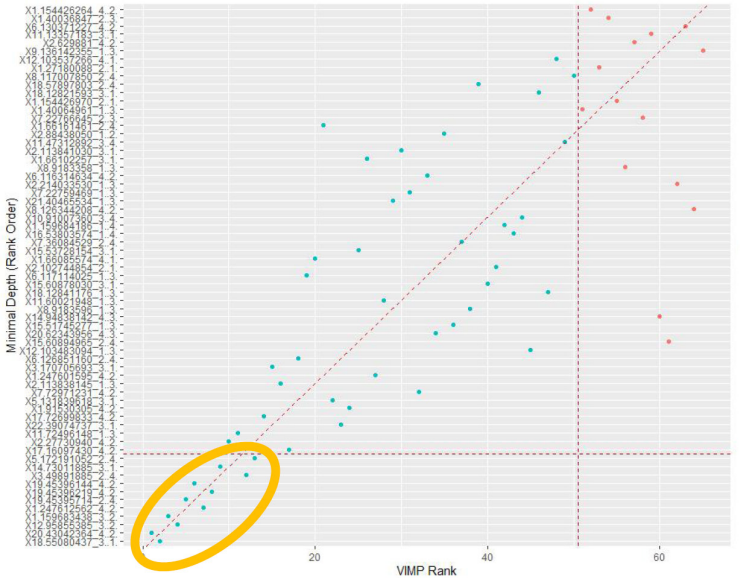


G

(1) 13 SNPs from our previous GWAS in WST ≤ 88

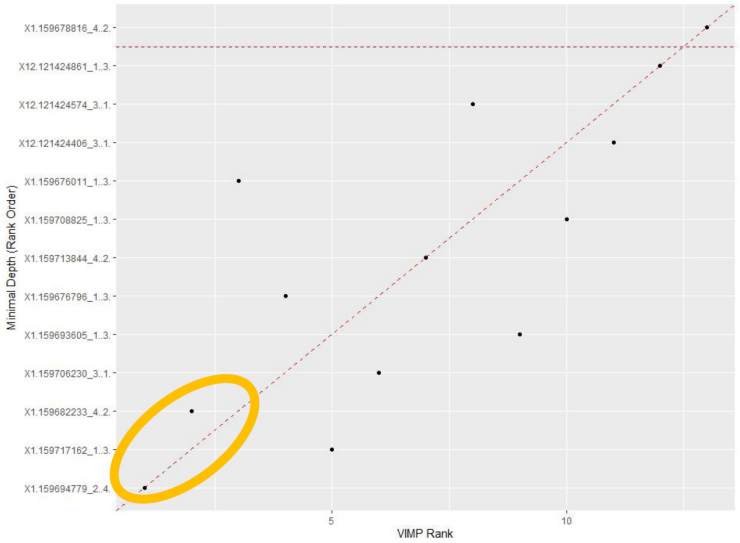


(2) 65 SNPs from outside GWASs in WST ≤ 88

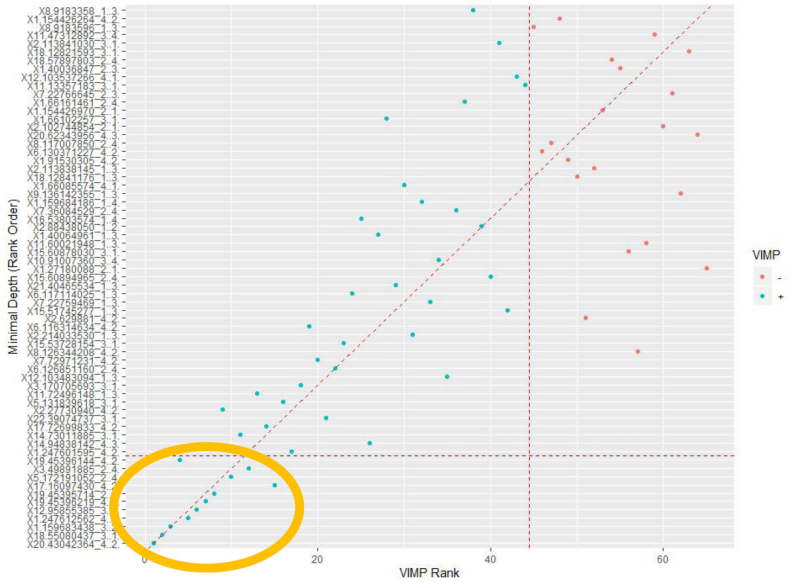


H

(1) 13 SNPs from our previous GWAS in WST > 88

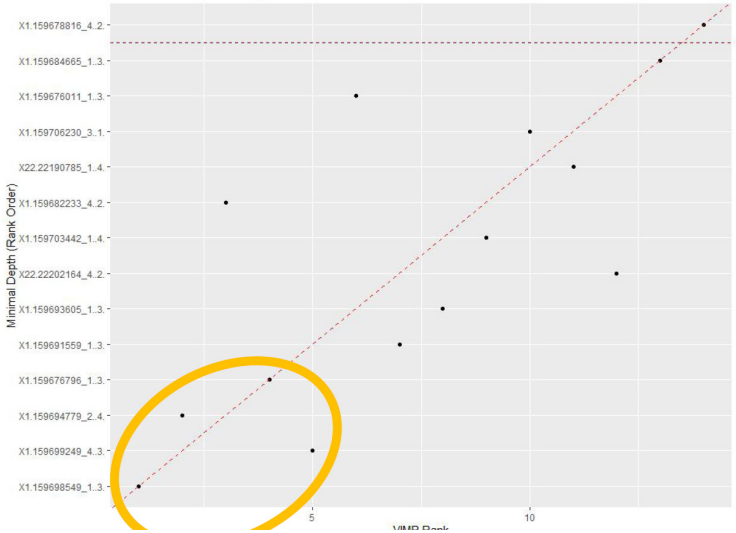


(2) 65 SNPs from outside GWASs in WST > 88

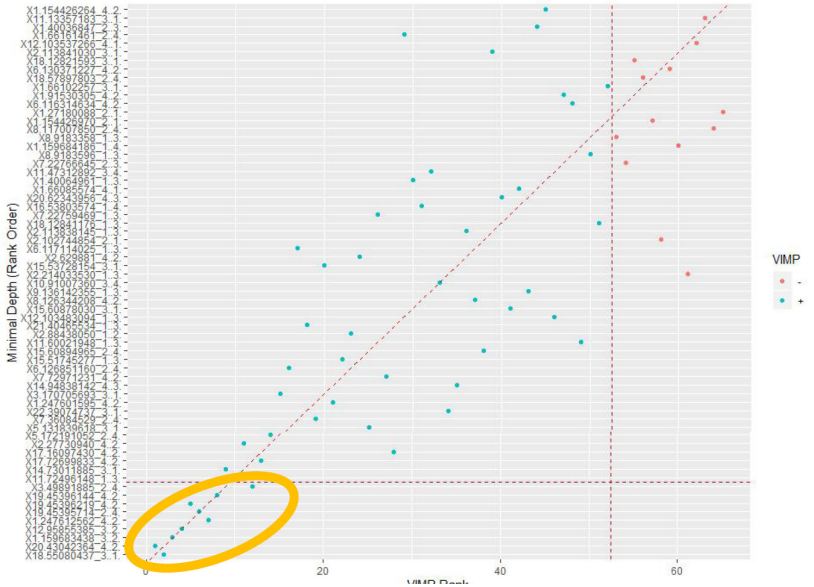


I

(1) 14 SNPs from our previous GWAS in MET ≥ 10

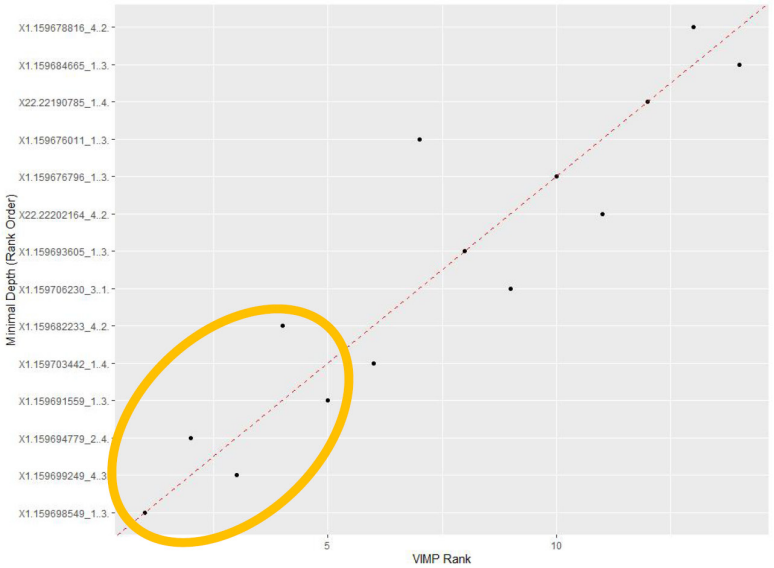


(2) 65 SNPs from outside GWASs in MET ≥ 10

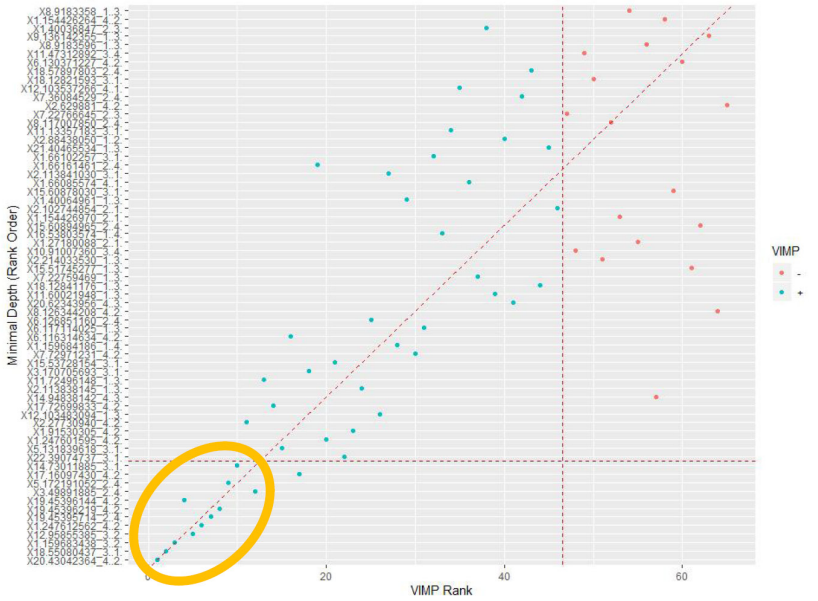


J

(1) 14 SNPs from our previous GWAS in MET < 10

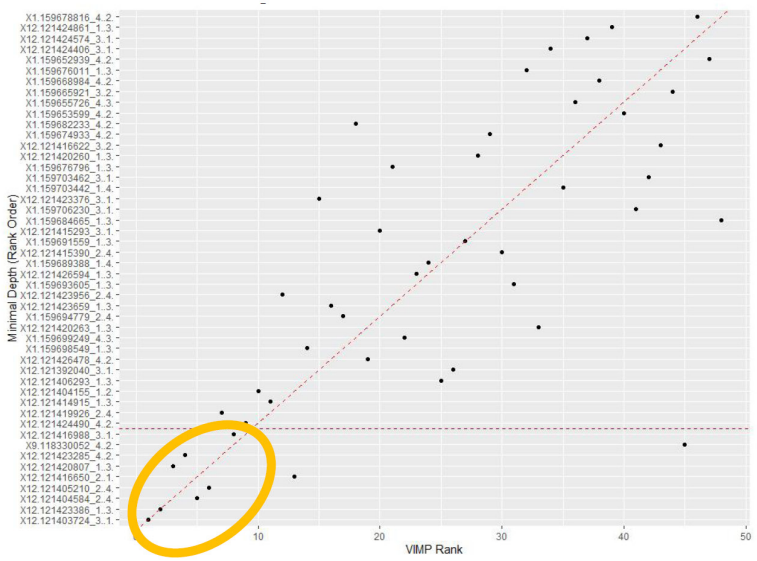


(2) 65 SNPs from outside GWASs in MET < 10

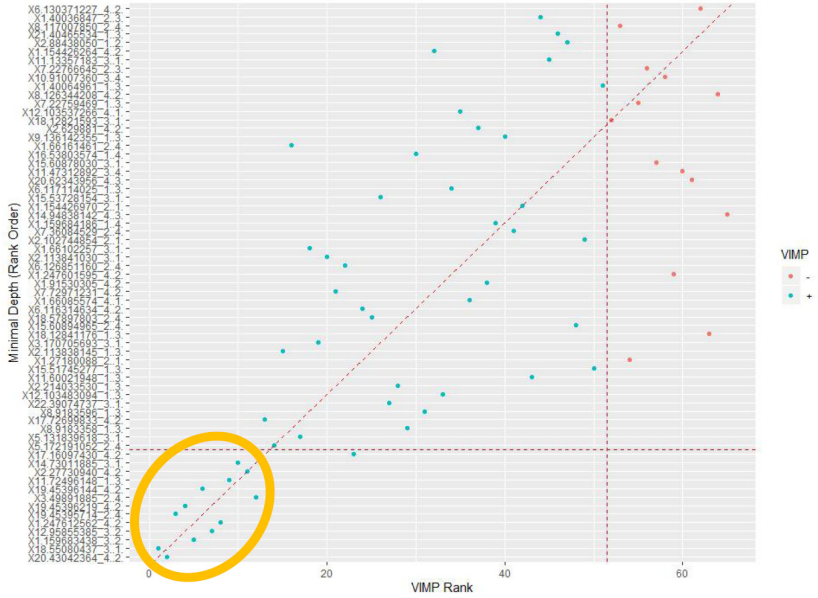


K

(1) 48 SNPs from our previous GWAS in SFA < 9.0



(2) 65 SNPs from outside GWASs in SFA < 9.0



L

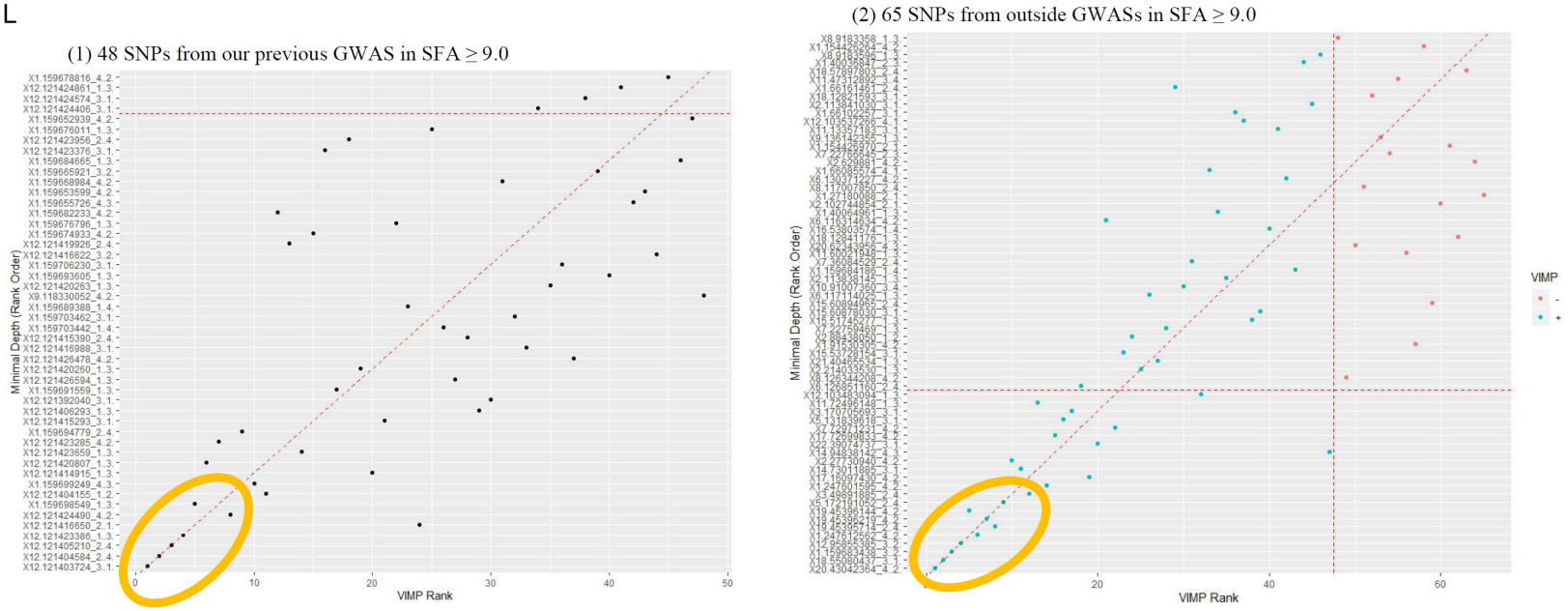


Figure S3. The first stage of random survival forest analysis, comparing minimal depth and VIMP rankings. A. Behavioral factors (BMI, body mass index; CVD, cardiovascular disease; E-only, exogenous estrogen; E+P, E + progestin; MFA, monounsaturated fatty acids; PFA, polyunsaturated fatty acids; SFA, saturated fatty acids; VIMP, variable of importance. Note: The 12 variables within the gold ellipse were identified as the most influential predictors). B. 152 (=87 + 65) SNPs in overall analysis (SNPs, single-nucleotide polymorphisms. Note: The 18 (=7 + 11) SNPs within the gold ellipses were identified as the most influential predictors). C. 114 (=49 + 65) SNPs in BMI-stratified analysis-non-overall obese group (BMI < 30) (BMI, body mass index. Note: The 18 (=9 + 9) SNPs within the gold ellipses were identified as the most influential predictors). D. 114 (=49 + 65) SNPs in BMI-stratified analysis - overall obese group (BMI ≥ 30) (Note: The 17 (=6 + 11) SNPs within the gold ellipses were identified as the most influential predictors). E. 67 SNPs (combining our 2 GWA SNPs and 65 SNPs from outside GWASs) in WHR-stratified analysis - non-viscerally obese group (WHR ≤ 0.85) (WHR, waist-to-hip ratio. Note: The 14 SNPs within the gold ellipse were identified as the most influential predictors). F. 67 SNPs (combining our 2 GWA SNPs and 65 SNPs from outside GWASs) in WHR-stratified analysis - viscerally obese group (WHR > 0.85) (Note: The 12 SNPs within the gold ellipse were identified as the most influential predictors). G. 78 (=13 + 65) SNPs in WST-stratified analysis - non-viscerally obese group (WST ≤ 88) (WST, waist circumference. Note: The 13 (=2 + 11) SNPs within the gold ellipses were identified as the most influential predictors). H. 78 (=13 + 65) SNPs in WST-stratified analysis-non-viscerally obese group (WST > 88) (Note: The 13 (=2 + 11) SNPs within the gold ellipses were identified as the most influential predictors). I. 79 (=14 + 65) SNPs in physical activity-stratified analysis - active group (MET ≥ 10) (MET, metabolic equivalent. Note: The 13 (=4 + 9) SNPs within the gold ellipses were identified as the most influential predictors). J. 79 (=14 + 65) SNPs in physical activity-stratified analysis-inactive group (MET < 10) (Note: The 16 (=5 + 11) SNPs within the gold ellipses were identified as the most influential predictors). K. 113 (=48 + 65) SNPs in SFA-stratified analysis- < 9.0% calories from SFA (SFA, saturated fatty acids. Note: The 19 (=7 + 12) SNPs within the gold ellipses were identified as the most influential predictors). L. 113 (=48 + 65) in SFA-stratified analysis - ≥ 9.0% calories from SFA (Note: The 16 (=6 + 10) SNPs within the gold ellipse were identified as the most influential predictors).

Table S2. GMDR-based model for high-order gene-gene interactions in relation to CRC risk

n	Model	TBA	*P* Value	CVC
	<Overall>			
1	*ONECUT2* rs4092465	0.6954	0.0010	10/10
2	*ONECUT2* rs4092465 *HNF4A* rs1800961	**0.7019**	**0.0010**	**10/10**
	<Non-overall obese group, BMI < 30 kg/m^2>			
1	*ONECUT2* rs4092465	0.6933	0.0010	10/10
2	*ONECUT2* rs4092465 *HNF4A* rs1800961	**0.6993**	**0.0010**	**10/10**
	<Overall obese group, BMI ≥ 30 kg/m^2>			
1	*ONECUT2* rs4092465	0.7006	0.0010	10/10
2	*ONECUT2* rs4092465 *HNF4A* rs1800961	**0.7084**	**0.0010**	**10/10**
	<Non-viscerally obese group, WHR ≤ 0.85>			
1	*ONECUT2* rs4092465	0.7010	0.0010	10/10
2	*ONECUT2* rs4092465 *HNF4A* rs1800961	**0.7106**	**0.0010**	**10/10**
	<Viscerally obese group, WHR > 0.85>			
1	*ONECUT2* rs4092465	**0.6819**	**0.0010**	**10/10**
2	*ONECUT2* rs4092465 *HNF4A* rs1800961	0.6769	0.0010	10/10
	<Non-viscerally obese group, WST ≤ 88 cm>			
1	*ONECUT2* rs4092465	0.6943	0.0010	10/10
2	*ONECUT2* rs4092465 *HNF4A* rs1800961	**0.7014**	**0.0010**	**10/10**
	<Viscerally obese group, WST > 88 cm>			
1	*ONECUT2* rs4092465	0.6970	0.0010	10/10
2	*ONECUT2* rs4092465 *HNF4A* rs1800961	**0.7029**	**0.0010**	**10/10**
	<Active group, MET ≥ 10.0>			
1	*ONECUT2* rs4092465	0.6925	0.0010	10/10
2	*ONECUT2* rs4092465 *HNF4A* rs1800961	**0.7014**	**0.0010**	**10/10**
	<Inactive group, MET < 10.0>			
1	*ONECUT2* rs4092465	0.6973	0.0010	10/10
2	*ONECUT2* rs4092465 *HNF4A* rs1800961	**0.7009**	**0.0010**	**10/10**
	<Low-fat diet group, % cal. from SFA < 9.0>			
1	*ONECUT2* rs4092465	0.7061	0.0010	10/10
2	*ONECUT2* rs4092465 *HNF4A* rs1800961	**0.7135**	**0.0010**	**10/10**
3	*ONECUT2* rs4092465 *HNF4A* rs1800961 *CRP* rs1800947	0.7094	0.0010	10/10
	<High-fat diet group, % cal. from SFA ≥ 9.0>			
1	*ONECUT2* rs4092465	0.6922	0.0010	10/10
2	*ONECUT2* rs4092465 *HNF4A* rs1800961	**0.6984**	**0.0010**	**10/10**

BMI, body mass index; CRC, colorectal cancer; CVC, cross-validation consistency; GMDR, generalized multifactor dimensionality reduction; MET, metabolic equivalent; SFA, saturated fatty acids; TBA, testing balance accuracy; WHR, waist-to-hip ratio; WST, waist circumference. Models in bold face are considered the best, with the highest TBA, 10/10 CVC, and $p < 0.05$.

Table S3A. The second stage of random survival forest analysis: predictive value of variables in non-overall obese group (BMI < 30 kg/m^2)

Variable*	Minimal Depth†	VIMP	C-index	Error¶	Drop Error§
Duration of oral contraceptive use	1.8458	0.0472	0.8617	0.1383	0.3617
HNF4A rs1800961	2.7410	0.0259	0.9332	0.0668	0.0715
ONECUT2 rs4092465	2.8580	0.0171	0.9532	0.0468	0.0201
CRP rs1800947	3.0740	0.0117	0.9523	0.0477	-0.0010
METAP2 rs11108056	3.3444	0.0107	0.9556	0.0444	0.0034
NLRP3 rs10925027	3.6380	0.0076	0.9585	0.0415	0.0029
TOMM40 rs157581	4.0532	0.0120	0.9569	0.0431	-0.0016
TOMM40 rs157582	4.2028	0.0091	0.9572	0.0428	0.0004
TRAIP s2352975	4.4012	0.0027	0.9585	0.0415	0.0013
Age at enrollment	4.6002	0.0017	0.9597	0.0403	0.0011
TOMM40 rs11556505	4.8112	0.0113	0.9598	0.0402	0.0001
Waist circumference	5.2336	0.0001	0.9597	0.0403	-0.0001
Hip circumference	5.2418	-0.0001	0.9592	0.0408	-0.0005
Waist-to-hip ratio	5.3210	0.0001	0.9593	0.0407	0.0001
Age at menopause	5.3436	0.0003	0.9598	0.0402	0.0005
HNF1A rs1169288	5.3498	0.0018	0.9587	0.0413	-0.0012
Duration of E+P use	5.4476	0.0007	0.9606	0.0394	0.0019
HNF1A rs11065385	5.5488	0.0056	0.9602	0.0398	-0.0004
Total months of breastfeeding	5.6758	0.0003	0.9589	0.0411	-0.0012
Height	5.6878	-0.0002	0.9585	0.0415	-0.0004
HNF1A rs11065384	5.7466	0.0057	0.9583	0.0417	-0.0002
HNF1A-AS1 rs7135337	5.7714	0.0030	0.9575	0.0425	-0.0007
Weight	5.9938	-0.0001	0.9571	0.0429	-0.0004
HNF1A rs2243458	6.1380	0.0030	0.9565	0.0435	-0.0006
HNF1A-AS1 rs2251468	6.1412	0.0058	0.9558	0.0442	-0.0007
Education	6.3760	-0.0001	0.9572	0.0428	0.0014
HNF1A-AS1 rs7953249	6.9202	0.0065	0.9570	0.0430	-0.0002
HNF1A-AS1 rs10774579	7.2118	0.0032	0.9566	0.0434	-0.0004
HNF1A-AS1 rs1920792	7.2578	0.0035	0.9555	0.0445	-0.0011

BMI, body mass index; C-index, concordance index; E+P, exogenous estrogen + progestin; VIMP, variable of importance. *Variables are ordered by minimal depth. †Predictive value of variable was assessed via minimal depth in the nested random survival forest models. A lower value is likely to have a greater impact on prediction. ¶The incremental error rate of each variable was estimated 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 3rd error rate was estimated from the 3rd nested model (including the 1st, 2nd, and 3rd variables). §The drop error rate was estimated by the difference between the error rates from the nested models with a prior and the corresponding variable. For example, the drop error rate of the 2nd variable was estimated by the difference between the error rates from the 1nd and 2rd nested models. The error rate for the null model is set at 0.5; thus, the drop error rate for the 1st variable was obtained by subtracting the error rate (0.3617) from 0.5.

Table S3B. The second stage of random survival forest analysis: predictive value of variables in overall obese group (BMI ≥ 30 kg/m^2)

Variable*	Minimal Depth†	VIMP	C-index	Error¶	Drop Error§
Duration of oral contraceptive use	2.4382	0.0344	0.8249	0.1751	0.3249
HNF4A rs1800961	2.9428	0.0334	0.9024	0.0976	0.0775
CRP rs1800947	3.7300	0.0155	0.9022	0.0978	-0.0002
ONECUT2 rs4092465	3.8762	0.0181	0.9373	0.0627	0.0351
METAP2 rs11108056	4.1160	0.0093	0.9383	0.0617	0.0010
NLRP3 rs10925027	4.4502	0.0068	0.9409	0.0591	0.0026
TOMM40 rs157582	5.2110	0.0125	0.9408	0.0592	-0.0001
RGS6 rs2239222	5.3524	0.0035	0.9416	0.0584	0.0008
TOMM40 rs157581	5.3750	0.0103	0.9424	0.0576	0.0008
Education	5.5236	-0.0005	0.9431	0.0569	0.0007
TOMM40 rs11556505	5.5752	0.0136	0.9471	0.0529	0.0040
DUSP1 rs17658229	5.6148	0.0038	0.9513	0.0487	0.0042
CD300LF/RAB37 rs10512597	5.7284	0.0015	0.9514	0.0486	0.0001
Height	5.8142	-0.0001	0.9496	0.0504	-0.0019
Age at enrollment	6.0544	0.0002	0.9466	0.0534	-0.0030
Age at menopause	6.0706	0.0000	0.9456	0.0544	-0.0010
Hip circumference	6.0802	0.0006	0.9479	0.0521	0.0023
HNF1A rs11065385	6.2498	0.0033	0.9469	0.0531	-0.0010
Waist circumference	6.3752	-0.0003	0.9462	0.0538	-0.0008
HNF1A-AS1 rs7135337	6.3972	0.0030	0.9447	0.0553	-0.0014
Waist-to-hip ratio	6.4432	-0.0001	0.9427	0.0573	-0.0021
HNF1A-AS1 rs7953249	6.4710	0.0080	0.9417	0.0583	-0.0009
HNF1A-AS1 rs2251468	6.5062	0.0061	0.9391	0.0609	-0.0027
Weight	6.7850	-0.0006	0.9381	0.0619	-0.0010
Duration of E+P use	6.9146	0.0000	0.9381	0.0619	0.0000
HNF1A-AS1 rs1920792	7.8064	0.0026	0.9394	0.0606	0.0013
HNF1A-AS1 rs10774579	7.9144	0.0026	0.9402	0.0598	0.0008
Total months of breastfeeding	8.0044	0.0000	0.9392	0.0608	-0.0010

BMI, body mass index; C-index, concordance index; E+P, exogenous estrogen + progestin; VIMP, variable of importance. *Variables are ordered by minimal depth. †Predictive value of variable was assessed via minimal depth in the nested random survival forest models. A lower value is likely to have a greater impact on prediction. ¶The incremental error rate of each variable was estimated 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 3rd error rate was estimated from the 3rd nested model (including the 1st, 2nd, and 3rd variables). §The drop error rate was estimated by the difference between the error rates from the nested models with a prior and the corresponding variable. For example, the drop error rate of the 2nd variable was estimated by the difference between the error rates from the 1nd and 2rd nested models. The error rate for the null model is set at 0.5; thus, the drop error rate for the 1st variable was obtained by subtracting the error rate (0.3249) from 0.5.

Table S3C. The second stage of random survival forest analysis: predictive value of variables in non-viscerally obese group (waist-to-hip ratio ≤ 0.85)

Variable*	Minimal Depth†	VIMP	C-index	Error¶	Drop Error§
Duration of oral contraceptive use	1.9340	0.0409	0.8510	0.1490	0.3510
HNF4A rs1800961	2.6618	0.0267	0.9326	0.0674	0.0817
ONECUT2 rs4092465	2.7684	0.0185	0.9540	0.0460	0.0214
CRP rs1800947	3.1084	0.0124	0.9519	0.0481	-0.0022
METAP2 rs11108056	3.3028	0.0101	0.9520	0.0480	0.0001
NLRP3 rs10925027	3.5750	0.0075	0.9551	0.0449	0.0032
TOMM40 rs157581	3.7964	0.0120	0.9556	0.0444	0.0005
TOMM40 rs157582	3.9188	0.0107	0.9562	0.0438	0.0006
TOMM40 rs11556505	4.3290	0.0132	0.9568	0.0432	0.0006
TRAIP rs2352975	4.3390	0.0029	0.9597	0.0403	0.0030
DUSP1 rs17658229	4.4170	0.0038	0.9646	0.0354	0.0049
Age at enrollment	4.6750	0.0010	0.9649	0.0351	0.0003
NCOR1 rs178810	4.8928	0.0003	0.9640	0.0360	-0.0010
RGS6 rs2239222	4.9694	0.0020	0.9627	0.0373	-0.0013
Duration of E+P use	5.0122	0.0011	0.9640	0.0360	0.0013
Hip circumference	5.0998	-0.0002	0.9639	0.0361	-0.0002
GCKR rs1260326	5.1030	0.0029	0.9649	0.0351	0.0011
Age at menopause	5.1118	0.0004	0.9641	0.0359	-0.0008
Waist circumference	5.1856	0.0000	0.9642	0.0358	0.0002
STARD10 rs7121935	5.2054	0.0025	0.9629	0.0371	-0.0014
Height	5.5332	-0.0002	0.9629	0.0371	0.0000
BMI	5.5796	-0.0002	0.9627	0.0373	-0.0002
Weight	5.6600	0.0001	0.9614	0.0386	-0.0013
Education	5.9200	-0.0003	0.9605	0.0395	-0.0009
Total months of breastfeeding	6.6006	0.0000	0.9602	0.0398	-0.0004

BMI, body mass index; C-index, concordance index; E+P, exogenous estrogen + progestin; VIMP, variable of importance. *Variables are ordered by minimal depth. †Predictive value of variable was assessed via minimal depth in the nested random survival forest models. A lower value is likely to have a greater impact on prediction. ¶The incremental error rate of each variable was estimated 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 3rd error rate was estimated from the 3rd nested model (including the 1st, 2nd, and 3rd variables). §The drop error rate was estimated by the difference between the error rates from the nested models with a prior and the corresponding variable. For example, the drop error rate of the 2nd variable was estimated by the difference between the error rates from the 1nd and 2rd nested models. The error rate for the null model is set at 0.5; thus, the drop error rate for the 1st variable was obtained by subtracting the error rate (0.3510) from 0.5.

Table S3D. The second stage of random survival forest analysis: predictive value of variables in viscerally obese group (waist-to-hip ratio > 0.85)

Variable*	Minimal Depth†	VIMP	C-index	Error¶	Drop Error§
Duration of oral contraceptive use	2.1058	0.0563	0.8342	0.1658	0.3342
HNF4A rs1800961	2.9236	0.0282	0.9160	0.0840	0.0818
CRP rs1800947	3.0800	0.0157	0.9134	0.0866	-0.0026
ONECUT2 rs4092465	3.4690	0.0226	0.9385	0.0615	0.0251
METAP2 rs11108056	4.0050	0.0085	0.9387	0.0613	0.0001
NLRP3 rs10925027	4.1224	0.0082	0.9379	0.0621	-0.0008
TOMM40 rs157581	4.4910	0.0216	0.9361	0.0639	-0.0018
TOMM40 rs11556505	4.5374	0.0255	0.9411	0.0589	0.0050
TOMM40 rs157582	4.8506	0.0176	0.9399	0.0601	-0.0012
Age at enrollment	4.9918	0.0030	0.9435	0.0565	0.0037
DUSP1 rs17658229	5.0486	0.0038	0.9486	0.0514	0.0051
TRAIP rs2352975	5.1480	0.0017	0.9493	0.0507	0.0007
NLRP3 rs12239046	5.2114	0.0036	0.9497	0.0503	0.0004
NCOR1 rs178810	5.3192	0.0009	0.9482	0.0518	-0.0015
Education	5.6916	-0.0006	0.9466	0.0534	-0.0017
Height	5.8588	-0.0001	0.9447	0.0553	-0.0019
Weight	6.0178	-0.0003	0.9455	0.0545	0.0008
Total months of breastfeeding	6.0396	0.0007	0.9456	0.0544	0.0001
BMI	6.0644	-0.0002	0.9441	0.0559	-0.0015
Hip circumference	6.2248	0.0002	0.9432	0.0568	-0.0010
Age at menopause	6.2702	-0.0005	0.9439	0.0561	0.0007
Waist circumference	6.3782	0.0004	0.9418	0.0582	-0.0021
Duration of E+P use	7.3304	-0.0001	0.9408	0.0592	-0.0010

BMI, body mass index; C-index, concordance index; E+P, exogenous estrogen + progestin; VIMP, variable of importance. *Variables are ordered by minimal depth. †Predictive value of variable was assessed via minimal depth in the nested random survival forest models. A lower value is likely to have a greater impact on prediction. ¶The incremental error rate of each variable was estimated 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 3rd error rate was estimated from the 3rd nested model (including the 1st, 2nd, and 3rd variables). §The drop error rate was estimated by the difference between the error rates from the nested models with a prior and the corresponding variable. For example, the drop error rate of the 2nd variable was estimated by the difference between the error rates from the 1nd and 2rd nested models. The error rate for the null model is set at 0.5; thus, the drop error rate for the 1st variable was obtained by subtracting the error rate (0.4251) from 0.5.

Table S3E. The second stage of random survival forest analysis: predictive value of variables in non-viscerally obese group (waist circumference ≤ 88 cm)

Variable*	Minimal Depth†	VIMP	C-index	Error¶	Drop Error§
Duration of oral contraceptive use	1.8380	0.0508	0.8749	0.1251	0.3749
ONECUT2 rs4092465	2.7844	0.0209	0.9479	0.0521	0.0730
HNF4A rs1800961	2.8618	0.0234	0.9534	0.0466	0.0055
CRP rs1800947	3.1218	0.0164	0.9534	0.0466	-0.0001
METAP2 rs11108056	3.2872	0.0094	0.9537	0.0463	0.0004
NLRP3 rs10925027	3.5948	0.0069	0.9560	0.0440	0.0022
TOMM40 rs157581	3.8174	0.0162	0.9554	0.0446	-0.0006
TOMM40 rs157582	3.9572	0.0136	0.9569	0.0431	0.0015
TOMM40 rs11556505	4.2854	0.0178	0.9568	0.0432	-0.0001
Age at enrollment	4.4032	0.0015	0.9580	0.0420	0.0012
TRAIP rs2352975	4.6390	0.0025	0.9592	0.0408	0.0013
RGS6 rs2239222	4.8606	0.0033	0.9592	0.0408	-0.0001
DUSP1 rs17658229	4.8922	0.0034	0.9623	0.0377	0.0031
Hip circumference	5.3536	-0.0002	0.9612	0.0388	-0.0011
Duration of E+P use	5.3606	0.0008	0.9617	0.0383	0.0005
Age at menopause	5.3786	0.0001	0.9614	0.0386	-0.0004
Waist-to-hip ratio	5.5536	-0.0002	0.9620	0.0380	0.0006
Height	5.5628	-0.0003	0.9616	0.0384	-0.0004
CRP rs7551731	5.6326	0.0013	0.9611	0.0389	-0.0005
Education	5.7548	-0.0002	0.9602	0.0398	-0.0008
BMI	5.8016	-0.0001	0.9600	0.0400	-0.0003
Weight	5.8430	0.0000	0.9588	0.0412	-0.0011
Total months of breastfeeding	5.9848	0.0001	0.9581	0.0419	-0.0007
CRP rs1205	6.4662	0.0012	0.9576	0.0424	-0.0005

BMI, body mass index; C-index, concordance index; E+P, exogenous estrogen + progestin; VIMP, variable of importance. *Variables are ordered by minimal depth. †Predictive value of variable was assessed via minimal depth in the nested random survival forest models. A lower value is likely to have a greater impact on prediction. ¶The incremental error rate of each variable was estimated 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 3rd error rate was estimated from the 3rd nested model (including the 1st, 2nd, and 3rd variables). §The drop error rate was estimated by the difference between the error rates from the nested models with a prior and the corresponding variable. For example, the drop error rate of the 2nd variable was estimated by the difference between the error rates from the 1nd and 2rd nested models. The error rate for the null model is set at 0.5; thus, the drop error rate for the 1st variable was obtained by subtracting the error rate (0.3749) from 0.5.

Table S3F. The second stage of random survival forest analysis: predictive value of variables in viscerally obese group (waist circumference > 88 cm)

Variable*	Minimal Depth†	VIMP	C-index	Error¶	Drop Error§
Duration of oral contraceptive use	2.1864	0.0392	0.8140	0.1860	0.3140
HNF4A rs1800961	2.5850	0.0300	0.9117	0.0883	0.0977
ONECUT2 rs4092465	3.1096	0.0206	0.9442	0.0558	0.0325
CRP rs1800947	3.1376	0.0203	0.9410	0.0590	-0.0032
METAP2 rs11108056	3.7004	0.0090	0.9425	0.0575	0.0015
NLRP3 rs10925027	3.7572	0.0086	0.9442	0.0558	0.0017
TOMM40 rs157581	4.2294	0.0170	0.9423	0.0577	-0.0019
TOMM40 rs11556505	4.3040	0.0207	0.9483	0.0517	0.0061
TOMM40 rs157582	4.3952	0.0149	0.9473	0.0527	-0.0010
TRAIP rs2352975	4.5670	0.0029	0.9500	0.0500	0.0027
DUSP1 rs17658229	4.5686	0.0048	0.9560	0.0440	0.0060
NCOR1 rs178810	4.7188	0.0010	0.9551	0.0449	-0.0009
Waist-to-hip ratio	5.2952	0.0008	0.9541	0.0459	-0.0010
Age at enrollment	5.3224	0.0011	0.9525	0.0475	-0.0016
Hip circumference	5.4536	0.0000	0.9512	0.0488	-0.0013
Weight	5.4926	-0.0001	0.9489	0.0511	-0.0023
CRP rs7551731	5.4994	0.0030	0.9506	0.0494	0.0017
BMI	5.5298	0.0001	0.9492	0.0508	-0.0014
Height	5.6658	0.0003	0.9483	0.0517	-0.0009
Age at menopause	5.6948	0.0003	0.9481	0.0519	-0.0003
Education	5.8200	-0.0003	0.9469	0.0531	-0.0012
Duration of E+P use	6.0930	0.0004	0.9473	0.0527	0.0004
CRP rs1205	6.5412	0.0022	0.9465	0.0535	-0.0009
Total months of breastfeeding	6.7822	-0.0001	0.9452	0.0548	-0.0013

BMI, body mass index; C-index, concordance index; E+P, exogenous estrogen + progestin; VIMP, variable of importance. *Variables are ordered by minimal depth. †Predictive value of variable was assessed via minimal depth in the nested random survival forest models. A lower value is likely to have a greater impact on prediction. ¶The incremental error rate of each variable was estimated 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 3rd error rate was estimated from the 3rd nested model (including the 1st, 2nd, and 3rd variables). §The drop error rate was estimated by the difference between the error rates from the nested models with a prior and the corresponding variable. For example, the drop error rate of the 2nd variable was estimated by the difference between the error rates from the 1nd and 2rd nested models. The error rate for the null model is set at 0.5; thus, the drop error rate for the 1st variable was obtained by subtracting the error rate (0.3140) from 0.5.

Table S3G. The second stage of random survival forest analysis: predictive value of variables in active group (MET ≥ 10.0)

Variable*	Minimal Depth†	VIMP	C-index	Error¶	Drop Error§
Duration of oral contraceptive use	1.9696	0.0472	0.8663	0.1337	0.3664
ONECUT2 rs4092465	2.8608	0.0232	0.9467	0.0533	0.0803
HNF4A rs1800961	3.0362	0.0250	0.9509	0.0491	0.0042
CRP rs1800947	3.3198	0.0181	0.9534	0.0466	0.0025
TOMM40 rs157581	3.6268	0.0192	0.9550	0.0450	0.0017
METAP2 rs11108056	3.6452	0.0099	0.9551	0.0449	0.0001
TOMM40 rs157582	3.6524	0.0182	0.9547	0.0453	-0.0004
NLRP3 rs10925027	3.9368	0.0073	0.9538	0.0462	-0.0009
TOMM40 rs11556505	4.1794	0.0188	0.9531	0.0469	-0.0007
Age at enrollment	4.6828	0.0020	0.9546	0.0454	0.0015
TRAIP rs2352975	4.7204	0.0021	0.9546	0.0454	0.0000
Waist-to-hip ratio	5.3104	0.0002	0.9542	0.0458	-0.0004
Hip circumference	5.5046	-0.0001	0.9530	0.0470	-0.0012
CRP rs4546916	5.6138	0.0018	0.9523	0.0477	-0.0006
CRP rs7553007	5.6618	0.0030	0.9516	0.0484	-0.0008
Duration of E+P use	5.7384	0.0005	0.9522	0.0478	0.0006
Age at menopause	5.8254	0.0005	0.9549	0.0451	0.0027
CRP rs7551731	5.8472	0.0024	0.9544	0.0456	-0.0005
Height	5.9668	0.0001	0.9539	0.0461	-0.0005
Waist circumference	6.0430	-0.0002	0.9531	0.0469	-0.0008
Weight	6.0468	0.0000	0.9520	0.0480	-0.0012
BMI	6.0648	0.0000	0.9515	0.0485	-0.0005
Education	6.3488	0.0002	0.9508	0.0492	-0.0007
Total months of breastfeeding	6.6838	0.0000	0.9507	0.0493	-0.0001
CRPP1 rs2808629	6.6970	0.0015	0.9497	0.0503	-0.0010

BMI, body mass index; C-index, concordance index; E+P, exogenous estrogen + progestin; VIMP, variable of importance. *Variables are ordered by minimal depth. †Predictive value of variable was assessed via minimal depth in the nested random survival forest models. A lower value is likely to have a greater impact on prediction. ¶The incremental error rate of each variable was estimated 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 3rd error rate was estimated from the 3rd nested model (including the 1st, 2nd, and 3rd variables). §The drop error rate was estimated by the difference between the error rates from the nested models with a prior and the corresponding variable. For example, the drop error rate of the 2nd variable was estimated by the difference between the error rates from the 1nd and 2rd nested models. The error rate for the null model is set at 0.5; thus, the drop error rate for the 1st variable was obtained by subtracting the error rate (0.3664) from 0.5.

Table S3H. The second stage of random survival forest analysis: predictive value of variables in inactive group (MET < 10.0)

Variable*	Minimal Depth†	VIMP	C-index	Error¶	Drop Error§
Duration of oral contraceptive use	1.9854	0.0442	0.8406	0.1594	0.3406
HNF4A rs1800961	2.6084	0.0284	0.9229	0.0771	0.0823
CRP rs1800947	2.9872	0.0218	0.9222	0.0778	-0.0007
ONECUT2 rs4092465	3.0140	0.0195	0.9514	0.0486	0.0292
METAP2 rs11108056	3.5080	0.0085	0.9487	0.0513	-0.0026
NLRP3 rs10925027	3.8052	0.0073	0.9510	0.0490	0.0023
TOMM40 rs11556505	4.2900	0.0202	0.9531	0.0469	0.0021
TOMM40 rs157581	4.3890	0.0164	0.9549	0.0451	0.0018
TOMM40 rs157582	4.4968	0.0126	0.9552	0.0448	0.0004
TRAIP rs2352975	4.7224	0.0024	0.9581	0.0419	0.0029
DUSP1 rs17658229	4.7960	0.0041	0.9640	0.0360	0.0059
Age at enrollment	4.8076	0.0012	0.9640	0.0360	0.0000
Age at menopause	5.3608	0.0000	0.9637	0.0363	-0.0003
RGS6 rs2239222	5.4870	0.0031	0.9622	0.0378	-0.0015
Height	5.5214	0.0002	0.9612	0.0388	-0.0010
Waist-to-hip ratio	5.5884	0.0000	0.9604	0.0396	-0.0009
Education	5.6460	-0.0003	0.9614	0.0386	0.0010
Hip circumference	5.7232	0.0001	0.9607	0.0393	-0.0007
Waist circumference	5.7434	0.0001	0.9603	0.0397	-0.0004
Weight	6.1228	0.0000	0.9594	0.0406	-0.0009
Duration of E+P use	6.1626	0.0006	0.9599	0.0401	0.0005
CRP rs4546916	6.1910	0.0017	0.9587	0.0413	-0.0012
BMI	6.1978	0.0000	0.9581	0.0419	-0.0006
Total months of breastfeeding	6.2060	0.0002	0.9574	0.0426	-0.0007
CRP rs7553007	6.2664	0.0023	0.9562	0.0438	-0.0012
CRP rs7551731	6.4542	0.0021	0.9574	0.0426	0.0013
CRP rs1341665	6.8222	0.0016	0.9570	0.0430	-0.0004
CRP rs1205	8.0524	0.0012	0.9564	0.0436	-0.0007

BMI, body mass index; C-index, concordance index; E+P, exogenous estrogen + progestin; MET metabolic equivalent; VIMP, variable of importance. *Variables are ordered by minimal depth. †Predictive value of variable was assessed via minimal depth in the nested random survival forest models. A lower value is likely to have a greater impact on prediction. ¶The incremental error rate of each variable was estimated 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 3rd error rate was estimated from the 3rd nested model (including the 1st, 2nd, and 3rd variables). §The drop error rate was estimated by the difference between the error rates from the nested models with a prior and the corresponding variable. For example, the drop error rate of the 2nd variable was estimated by the difference between the error rates from the 1nd and 2rd nested models. The error rate for the null model is set at 0.5; thus, the drop error rate for the 1st variable was obtained by subtracting the error rate (0.3406) from 0.5.

Table S3I. The second stage of random survival forest analysis: predictive value of variables in low fat-diet group (< 9.0% calories from SFA)

Variable*	Minimal Depth†	VIMP	C-index	Error¶	Drop Error§
Duration of oral contraceptive use	2.3760	0.0380	0.8533	0.1467	0.3533
HNF4A rs1800961	3.4392	0.0255	0.9087	0.0913	0.0554
CRP rs1800947	3.7850	0.0138	0.9173	0.0827	0.0085
ONECUT2 rs4092465	3.9060	0.0224	0.9458	0.0542	0.0285
METAP2 rs11108056	4.4608	0.0085	0.9440	0.0560	-0.0019
NLRP3 rs10925027	4.5676	0.0059	0.9476	0.0524	0.0036
TOMM40 rs157581	4.8704	0.0147	0.9482	0.0518	0.0006
TOMM40 rs157582	4.9000	0.0127	0.9459	0.0541	-0.0023
Age at enrollment	5.2120	0.0030	0.9475	0.0525	0.0017
TRAIP rs2352975	5.2532	0.0036	0.9498	0.0502	0.0023
TOMM40 rs11556505	5.7586	0.0102	0.9513	0.0487	0.0015
GCKR rs1260326	5.8828	0.0034	0.9528	0.0472	0.0015
Total months of breastfeeding	5.9300	0.0004	0.9524	0.0476	-0.0004
STARD10 rs7121935	6.0306	0.0045	0.9509	0.0491	-0.0015
Waist-to-hip ratio	6.0872	0.0007	0.9508	0.0492	-0.0001
Duration of E+P use	6.4696	0.0008	0.9526	0.0474	0.0018
Height	6.6032	-0.0001	0.9552	0.0448	0.0026
RGS6 rs2239222	6.6070	0.0028	0.9539	0.0461	-0.0012
HNF1A rs11065385	6.6214	0.0065	0.9530	0.0470	-0.0009
HNF1A rs11065384	6.7248	0.0059	0.9517	0.0483	-0.0013
HNF1A rs1183910	6.7818	0.0062	0.9514	0.0486	-0.0003
Age at menopause	6.7838	0.0002	0.9513	0.0487	-0.0001
Hip circumference	6.8202	0.0002	0.9503	0.0497	-0.0010
BMI	6.8798	0.0003	0.9511	0.0489	0.0009
HNF1A rs2244608	6.9668	0.0038	0.9490	0.0510	-0.0021
Waist circumference	7.0358	-0.0001	0.9502	0.0498	0.0012
HNF1A-AS1 rs7953249	7.0712	0.0036	0.9507	0.0493	0.0005
Weight	7.1228	0.0002	0.9505	0.0495	-0.0002
Education	7.2266	-0.0003	0.9494	0.0506	-0.0010
HNF1A-AS1 rs10774579	7.3834	0.0023	0.9479	0.0521	-0.0015
HNF1A-AS1 rs1920792	7.4540	0.0023	0.9480	0.0520	0.0000

BMI, body mass index; C-index, concordance index; E+P, exogenous estrogen + progestin; SFA, saturated fatty acids; VIMP, variable of importance. *Variables are ordered by minimal depth. †Predictive value of variable was assessed via minimal depth in the nested random survival forest models. A lower value is likely to have a greater impact on prediction. ¶The incremental error rate of each variable was estimated 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 3rd error rate was estimated from the 3rd nested model (including the 1st, 2nd, and 3rd variables). §The drop error rate was estimated by the difference between the error rates from the nested models with a prior and the corresponding variable. For example, the drop error rate of the 2nd variable was estimated by the difference between the error rates from the 1nd and 2rd nested models. The error rate for the null model is set at 0.5; thus, the drop error rate for the 1st variable was obtained by subtracting the error rate (0.3533) from 0.5.

Table S3J. The second stage of random survival forest analysis: predictive value of variables in high fat-diet group (≥ 9.0% calories from SFA)

Variable*	Minimal Depth†	VIMP	C-index	Error¶	Drop Error§
Duration of oral contraceptive use	1.9438	0.0480	0.8444	0.1556	0.3444
HNF4A rs1800961	2.4968	0.0297	0.9318	0.0682	0.0874
ONECUT2 rs4092465	2.8286	0.0169	0.9561	0.0439	0.0243
CRP rs1800947	3.0134	0.0153	0.9552	0.0448	-0.0009
METAP2 rs11108056	3.2954	0.0102	0.9561	0.0439	0.0009
NLRP3 rs10925027	3.4476	0.0082	0.9572	0.0428	0.0011
TOMM40 rs157581	4.0982	0.0127	0.9559	0.0441	-0.0013
TOMM40 rs157582	4.1014	0.0115	0.9570	0.0430	0.0011
TOMM40 rs11556505	4.3028	0.0148	0.9579	0.0421	0.0009
TRAIP rs2352975	4.3790	0.0028	0.9604	0.0396	0.0025
DUSP1 rs17658229	4.4004	0.0053	0.9659	0.0341	0.0055
Age at enrollment	4.7984	0.0011	0.9656	0.0344	-0.0003
Waist-to-hip ratio	5.3352	0.0000	0.9653	0.0347	-0.0004
HNF1A rs11065385	5.3556	0.0055	0.9650	0.0350	-0.0003
Hip circumference	5.3852	0.0000	0.9640	0.0360	-0.0010
Age at menopause	5.3868	0.0001	0.9634	0.0366	-0.0006
Duration of E+P use	5.5778	0.0006	0.9649	0.0351	0.0015
CRP rs7553007	5.5802	0.0014	0.9647	0.0353	-0.0002
Education	5.6508	-0.0003	0.9643	0.0357	-0.0004
Height	5.6578	-0.0001	0.9631	0.0369	-0.0012
Waist circumference	5.7086	-0.0001	0.9624	0.0376	-0.0006
BMI	5.9430	-0.0003	0.9616	0.0384	-0.0008
HNF1A-AS1 rs7953249	5.9834	0.0045	0.9615	0.0385	-0.0001
HNF1A rs2243458	6.0142	0.0027	0.9604	0.0396	-0.0011
Weight	6.0170	-0.0001	0.9597	0.0403	-0.0008
Total months of breastfeeding	6.5938	0.0001	0.9615	0.0385	0.0019
HNF1A-AS1 rs10774579	6.9126	0.0043	0.9608	0.0392	-0.0007
HNF1A-AS1 rs1920792	6.9468	0.0044	0.9603	0.0397	-0.0005

BMI, body mass index; C-index, concordance index; E+P, exogenous estrogen + progestin; SFA, saturated fatty acids; VIMP, variable of importance. *Variables are ordered by minimal depth. †Predictive value of variable was assessed via minimal depth in the nested random survival forest models. A lower value is likely to have a greater impact on prediction. ¶The incremental error rate of each variable was estimated 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 3rd error rate was estimated from the 3rd nested model (including the 1st, 2nd, and 3rd variables). §The drop error rate was estimated by the difference between the error rates from the nested models with a prior and the corresponding variable. For example, the drop error rate of the 2nd variable was estimated by the difference between the error rates from the 1nd and 2rd nested models. The error rate for the null model is set at 0.5; thus, the drop error rate for the 1st variable was obtained by subtracting the error rate (0.3444) from 0.5.

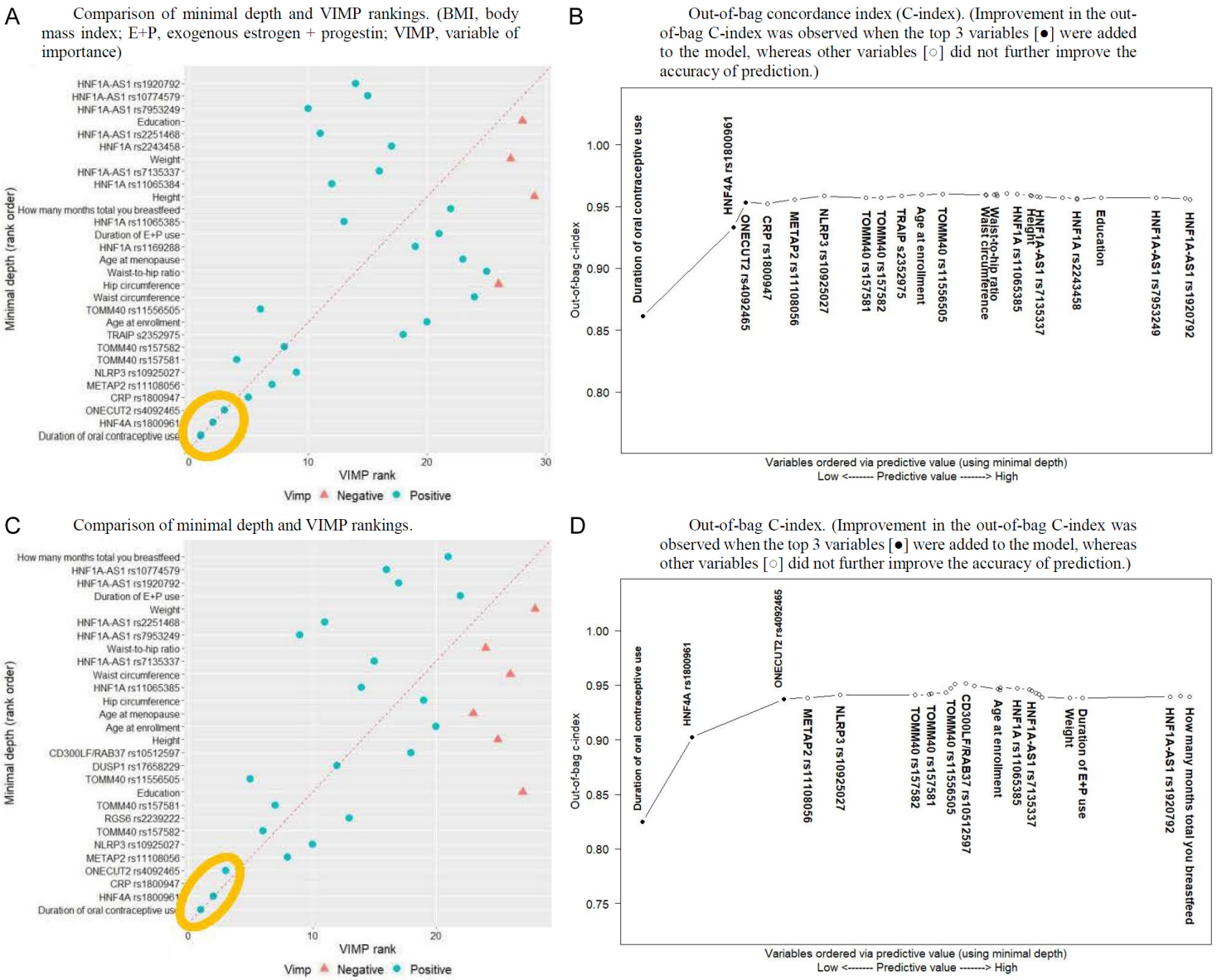


Figure S4. BMI-stratified analysis: the second stage of random survival forest (RSF) with 18 single-nucleotide polymorphisms (SNPs) and 12 behavioral factors selected from the first stage of RSF in non-overall obese group (BMI < 30, A and B); and with 17 SNPs and 12 behavioral factors selected in overall obese group (BMI ≥ 30, C and D). The variables within the gold ellipses in (A and C) were identified as the most influential predictors.

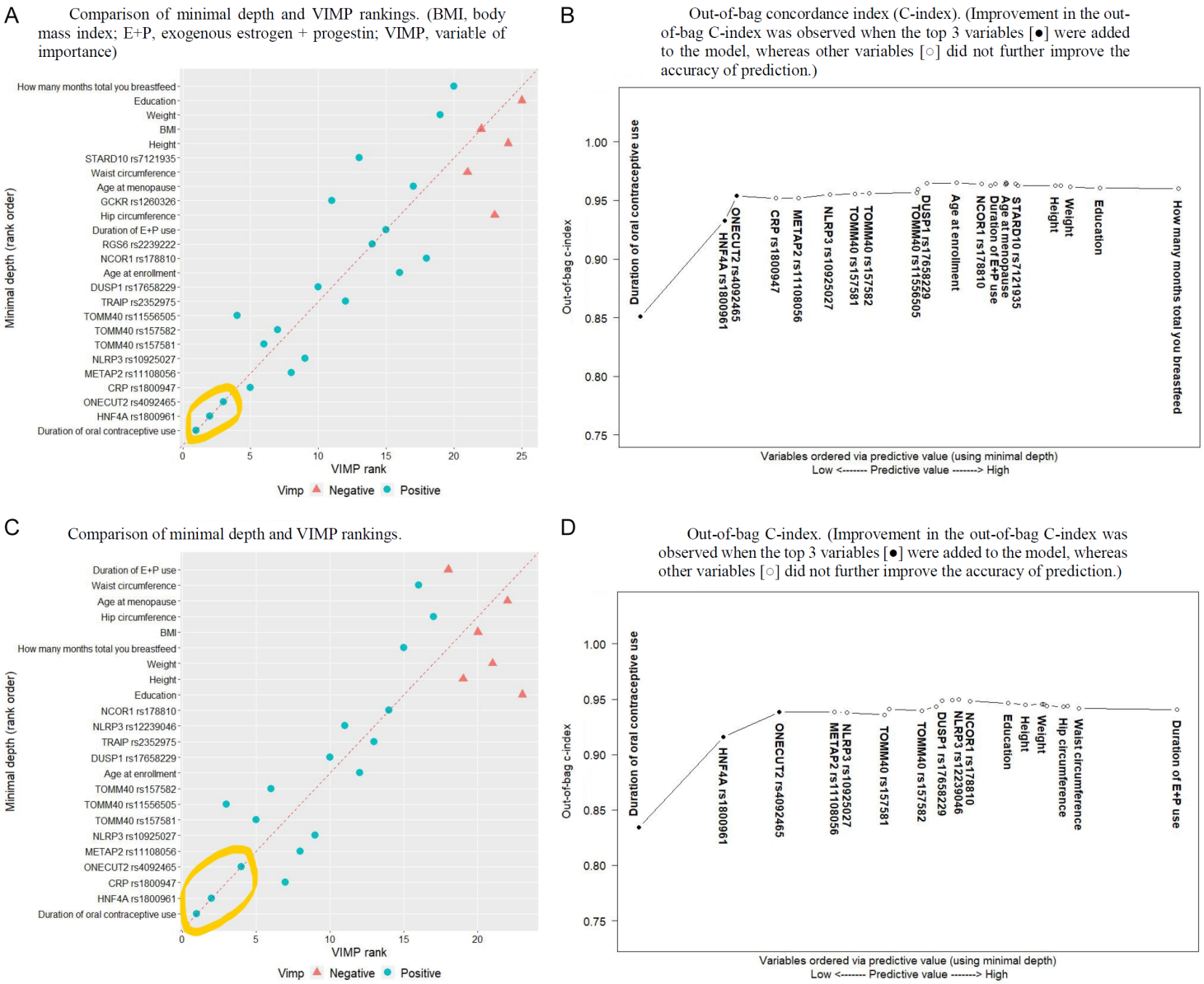


Figure S5. WHR-stratified analysis: the second stage of random survival forest (RSF) with 14 single-nucleotide polymorphisms (SNPs) and 12 behavioral factors selected from the first stage of RSF in non-viscerally obese group (WHR ≤ 0.85, A and B); and with 12 SNPs and 12 behavioral factors selected in viscerally obese group (WHR > 0.85, C and D). The variables within the gold ellipses in (A and C) were identified as the most influential predictors. (WHR, waist-to-hip ratio).

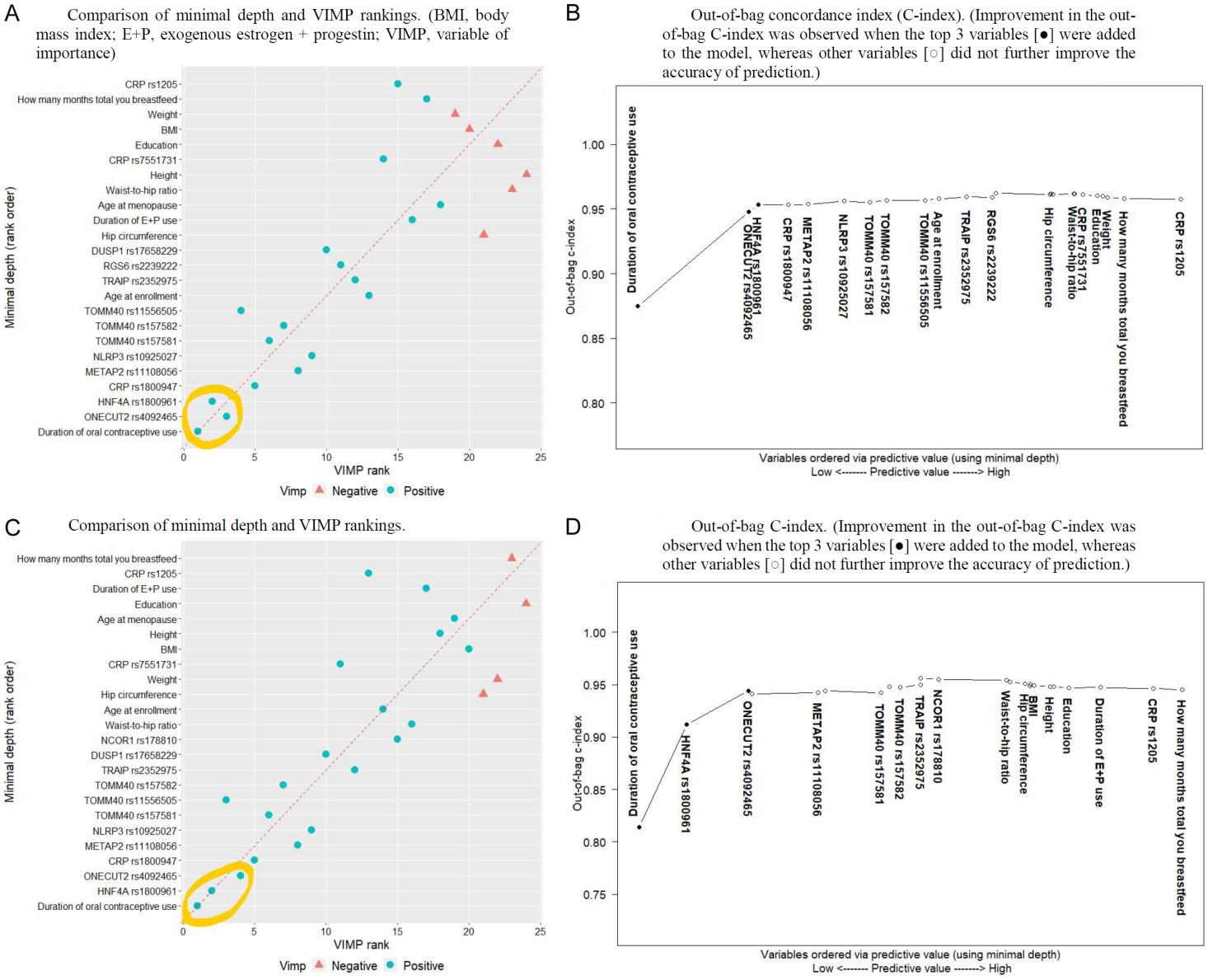


Figure S6. WST-stratified analysis: the second stage of random survival forest (RSF) with 13 single-nucleotide polymorphisms (SNPs) and 12 behavioral factors selected from the first stage of RSF in non-viscerally obese group (WST ≤ 88, A and B); and with 13 SNPs and 12 behavioral factors in viscerally obese group (WST > 88, C and D). The variables within the gold ellipses in (A and C) were identified as the most influential predictors. (WST, waist circumference.).

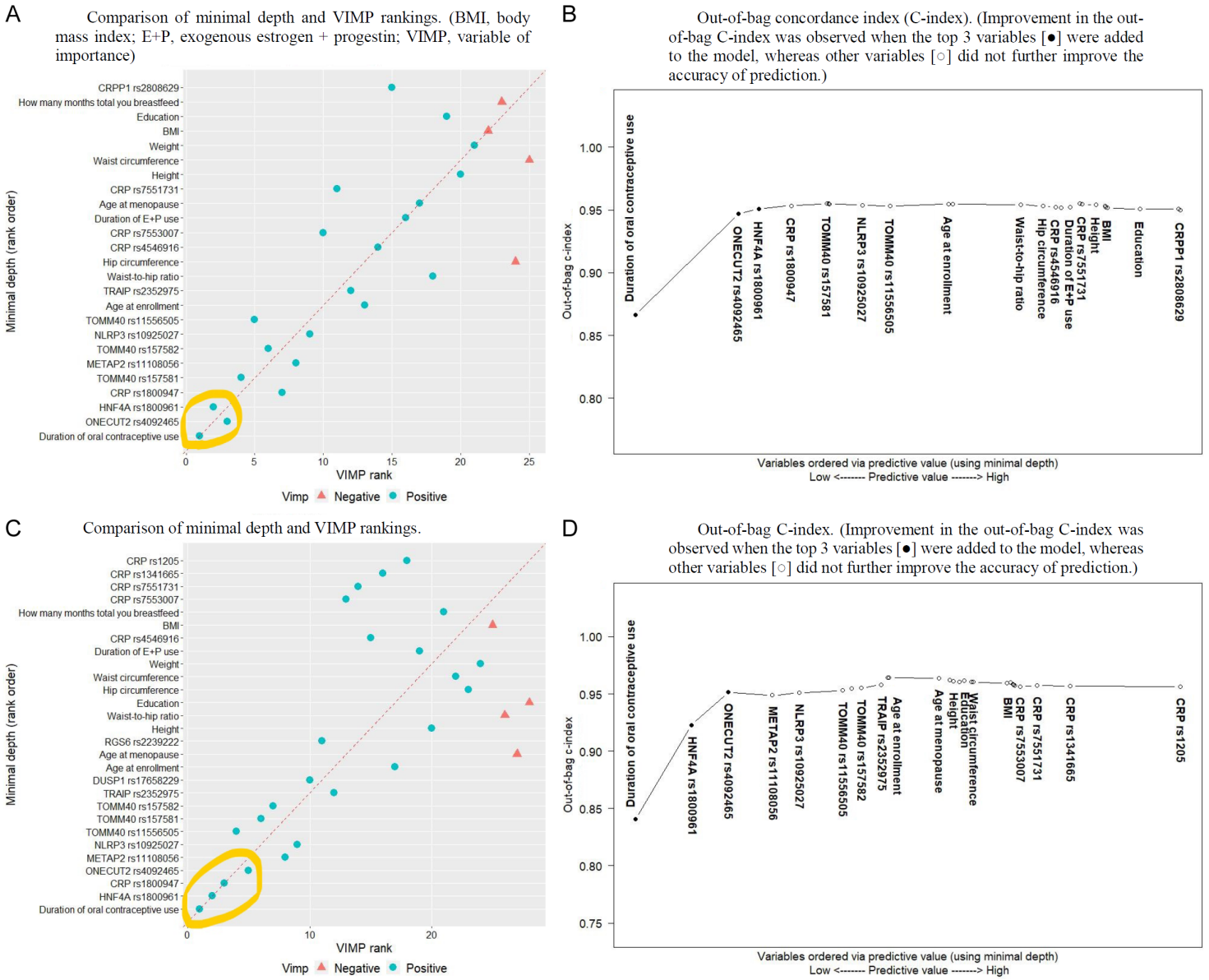


Figure S7. Physical activity-stratified analysis: the second stage of random survival forest (RSF) with 13 single-nucleotide polymorphisms (SNPs) and 12 behavioral factors selected from the first stage of RSF in active group (metabolic equivalent [MET] ≥ 10, A and B); and with 16 SNPs and 12 behavioral factors in inactive group (MET < 10, C and D). The variables within the gold ellipses in (A and C) were identified as the most influential predictors.

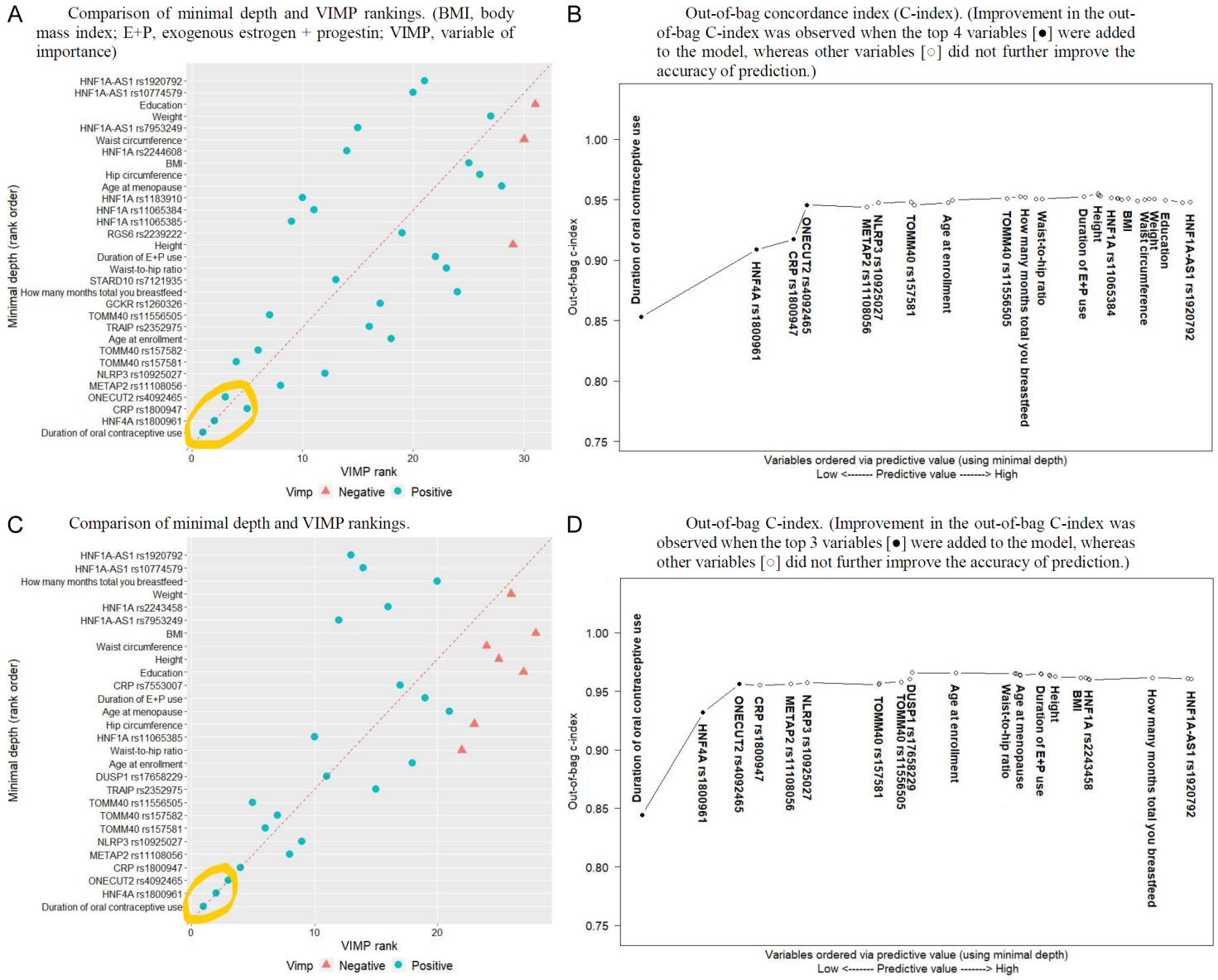


Figure S8. SFA-stratified analysis: the second stage of random survival forest (RSF) with 19 single-nucleotide polymorphisms (SNPs) and 12 behavioral factors selected from the first stage of RSF in low fat-diet group (< 9.0% calories from SFA, A and B); and with 16 SNPs and 12 behavioral factors in high fat-diet group (≥ 9.0% calories from SFA, C and D). The variables within the gold ellipses in (A and C) were identified as the most influential predictors. (SFA, saturated fatty acids.)

Table S4A. Overall analysis: results from multivariate regression predicting CRC risk

Variable	HR† (95% CI)	*p*
SNP (Ref/Alt)		
ONECUT2 rs4092465 (GG + AA/GA)	**7.49 (5.95-9.42)**	**< 2e-16**
HNF4A rs1800961 (CC + CT/TT)	**2.68 (1.61-4.48)**	**0.000158**
Behavioral factor*		
Duration of oral contraceptive use	**2.88 (2.48-3.35)**	**< 2e-16**

Alt, alternative allele; CI, confidence interval; CRC, colorectal cancer; HR, hazard ratio; Ref, reference allele; SNP, single-nucleotide polymorphism. Numbers in bold face are statistically significant. †Multivariate regression was adjusted by age at enrollment, education, body mass index, waist-to-hip ratio, age at menopause, total months of breastfeeding, and exogenous estrogen plus progestin. *A behavioral factor was analyzed as a binary variable via a cutoff value, where the cutoff level and/or higher reflects a greater risk for CRC on the basis of random survival forest analysis. The cutoff for oral contraceptive use was 5.1 years.

Table S4B. Stratified analysis: results from multivariate regression predicting CRC risk

Variable	HR† (95% CI)	*p*
<Non-overall obese group (BMI < 30 kg/m^2) (n=7,151)>		
SNPs (Ref/Alt)		
ONECUT2 rs4092465 (GG + AA/GA)	**7.25 (5.56-9.47)**	**< 2e-16**
HNF4A rs1800961 (CC + CT/TT)	**2.63 (1.45-4.78)**	**0.00153**
Behavioral factor*		
Duration of oral contraceptive use	**3.01 (2.52-3.58)**	**< 2e-16**
<Overall obese group (BMI ≥ 30 kg/m^2) (n=2,991)>		
SNPs (Ref/Alt)		
ONECUT2 rs4092465 (GG + AA/GA)	**8.27 (5.26-13.01)**	**< 2e-16**
HNF4A rs1800961 (CC + CT/TT)	**2.90 (1.08-7.85)**	**0.03549**
Behavioral factor*		
Duration of oral contraceptive use	**2.76 (2.06-3.70)**	**8.91E-12**
<Non-viscerally obese group, WHR ≤ 0.85 (n=7,222)>		
SNPs (Ref/Alt)		
ONECUT2 rs4092465 (GG + AA/GA)	**8.27 (6.21-11.02)**	**< 2e-16**
HNF4A rs1800961 (CC + CT/TT)	**4.38 (1.96-9.81)**	**0.000321**
Behavioral factor*		
Duration of oral contraceptive use	**3.08 (2.57-3.68)**	**< 2e-16**
<Viscerally obese group, WHR > 0.85 (n=2,920)>		
SNPs (Ref/Alt)		
ONECUT2 rs4092465 (GG + AA/GA)	**6.04 (4.11-8.88)**	**< 2e-16**
Behavioral factor*		
Duration of oral contraceptive use	**2.44 (1.85-3.21)**	**1.88E-10**
<Non-viscerally obese group, WST ≤ 88 cm (n=5,998)>		
SNPs (Ref/Alt)		
ONECUT2 rs4092465 (GG + AA/GA)	**7.53 (5.55-10.23)**	**< 2e-16**
HNF4A rs1800961 (CC + CT/TT)	**4.50 (1.86-10.88)**	**0.000835**
Behavioral factor*		
Duration of oral contraceptive use	**3.46 (2.84-4.21)**	**< 2e-16**
<Viscerally obese group, WST > 88 cm (n=4,144)>		
SNPs (Ref/Alt)		
ONECUT2 rs4092465 (GG + AA/GA)	**7.42 (5.24-10.50)**	**< 2e-16**
HNF4A rs1800961 (CC + CT/TT)	**1.79 (0.95-3.37)**	**0.06981**

Behavioral factor*		
Duration of oral contraceptive use	**2.38 (1.89-3.00)**	**1.28E-13**
<Active group, MET ≥ 10.0 (n=4,203)>		
SNPs (Ref/Alt)		
ONECUT2 rs4092465 (GG + AA/GA)	**7.42 (5.23-10.55)**	**< 2e-16**
HNF4A rs1800961 (CC + CT/TT)	**3.28 (1.35-7.95)**	**0.00850**
Behavioral factor*		
Duration of oral contraceptive use	**2.57 (2.04-3.22)**	**3.96E-16**
<Inactive group, MET < 10.0 (n=5,939)>		
SNPs (Ref/Alt)		
ONECUT2 rs4092465 (GG + AA/GA)	**7.49 (5.53-10.16)**	**< 2e-16**
HNF4A rs1800961 (CC + CT/TT)	**2.34 (1.25-4.39)**	**0.00792**
Behavioral factor*		
Duration of oral contraceptive use	**3.15 (2.58-3.86)**	**< 2e-16**
<Low fat-diet group, < 9.0% cal. from SFA (n=2,290)>		
SNPs (Ref/Alt)		
ONECUT2 rs4092465 (GG + AA/GA)	**9.91 (5.92-16.62)**	**< 2e-16**
HNF4A rs1800961 (CC + CT/TT)	**5.99 (1.48-24.26)**	**0.012135**
Behavioral factor*		
Duration of oral contraceptive use	**3.60 (2.63-4.93)**	**1.21E-15**
<High fat-diet group, ≥ 9.0% cal. from SFA (n=7,852)>		
SNPs (Ref/Alt)		
ONECUT2 rs4092465 (GG + AA/GA)	**7.01 (5.42-9.06)**	**< 2e-16**
HNF4A rs1800961 (CC + CT/TT)	**2.26 (1.30-3.92)**	**0.00368**
Behavioral factor*		
Duration of oral contraceptive use	**2.70 (2.27-3.20)**	**< 2e-16**

Alt, alternative allele; BMI, body mass index; CI, confidence interval; CRC, colorectal cancer; HR, hazard ratio; MET, metabolic equivalent; Ref, reference allele; SFA, saturated fatty acids; SNP, single-nucleotide polymorphism; WHR, waist-to-hip ratio; WST, waist circumference. Numbers in bold face are statistically significant. †Multivariate regression was adjusted by age at enrollment, education, BMI, WHR, age at menopause, total months of breastfeeding, and exogenous estrogen plus progestin; variables used to stratify were not included as covariates in the multivariate analysis; additional variables were adjusted in specific strata (height and weight were adjusted in BMI strata; waist and hip circumferences, in WHR strata; and hip circumference instead of WHR, in WST strata). *A behavioral factor was analyzed as a binary variable via a cutoff value, where the cutoff level and/or higher reflects a greater risk for CRC on the basis of random survival forest analysis. The cutoff for oral contraceptive use was 5.1 years.

Table S5. Combined effect of past OC use and risk genotypes (ONECUT2 rs4092465 GA; HNF4A rs1800961 TT) on CRC risk in obesity strata

n£	n total	HR† (95% CI)	*p**
<Non-viscerally obese group, WHR ≤ 0.85 (n=7,222)>			
0	2,590	reference	
1	3,525	**5.45 (3.78-7.85)**	**< 2e-16**
2	1,107	**22.63 (15.68-32.66)**	**< 2e-16**
p_{trend}			**< 2e-16**
<Viscerally obese group, WHR > 0.85 (n=2,920)>§			
0	908	reference	
1	1,456	**3.01 (1.88-4.83)**	**4.98E-06**
2	556	**11.00 (6.84-17.69)**	**< 2e-16**
p_{trend}			**< 2e-16**
<Non-viscerally obese group, WST ≤ 88 cm (n=5,998)>			
0	2,206	reference	
1	2,953	**4.95 (3.39-7.24)**	**< 2e-16**
2	839	**22.06 (15.04-32.35)**	**< 2e-16**
p_{trend}			**< 2e-16**
<Viscerally obese group, WST > 88 cm (n=4,144)>			
0	1,353	reference	
1	1,996	**3.78 (2.49-5.76)**	**5.35E-10**
2	795	**13.64 (8.96-20.76)**	**< 2e-16**
p_{trend}			**< 2e-16**
<Active group, MET ≥ 10.0 (n=4,203)>			
0	1,481	reference	
1	2,105	**4.71 (3.09-7.19)**	**6.12E-13**
2	617	**16.80 (10.92-25.85)**	**< 2e-16**
p_{trend}			**< 2e-16**
<Inactive group, MET < 10.0 (n=5,939)>			
0	2,078	reference	
1	2,844	**4.16 (2.85-6.07)**	**1.42E-13**
2	1,017	**18.03 (12.38-26.27)**	**< 2e-16**
p_{trend}			**< 2e-16**
<Low fat-diet group, < 9.0% cal. from SFA (n=2,290)>			
0	788	reference	
1	1,137	**3.36 (1.88-5.99)**	**4.05E-05**
2	365	**19.93 (11.28-35.21)**	**< 2e-16**
p_{trend}			**< 2e-16**
<High fat-diet group, ≥ 9.0% cal. from SFA (n=7,852)>			
0	2,771	reference	
1	3,812	**4.75 (3.44-6.56)**	**< 2e-16**
2	1,269	**16.94 (12.22-23.49)**	**< 2e-16**
p_{trend}			**< 2e-16**

CI, confidence interval; CRC, colorectal cancer; HR, hazard ratio; MET, metabolic equivalent; OC, oral contraceptive; SFA, saturated fatty acids; WHR, waist-to-hip ratio; WST, waist circumference. Numbers in bold face are statistically significant. £The combined number of risk genotypes and behavioral factors was based on 1) risk genotypes defined as 0 (low risk: none or 1 risk allele) and 1 (high risk: 2 risk alleles) and 2) behavioral factors defined as 0 (low risk: past OC use ≥ 5.1 years) and 1 (high risk: past OC use < 5.1 years). The ultimate number of combined risk genotypes and behavioral factors was defined as 0 (low risk for genotypes and behaviors), 1 (high risk for either genotypes or behaviors), and 2 (high risk for both genotypes and behaviors). §In the viscerally obese group (WHR), only the *ONECUT2* rs4092465 GA genotype was analyzed. †Multivariate regression was adjusted by age at enrollment, education, body mass index, hip and waist circumferences (both in WHR strata; hip only in WST strata), waist-to-hip ratio (in MET and SFA strata), age at menopause, total months of breastfeeding, and exogenous estrogen plus progestin. **p* values were adjusted to correct for multiple testing via the Benjamini-Hochberg approach.

Table S6. Joint effect of past OC use with risk genotypes (ONECUT2 rs4092465 GA; HNF4A rs1800961 TT) on CRC risk in obesity strata

n£	Total		Past OC use ≥ 5.1 years			Past OC use < 5.1 years		
	HR† (95% CI)	*p**	n	HR† (95% CI)	*p**	n	HR† (95% CI)	*p**
<Non-viscerally obese group, WHR ≤ 0.85 (n=7,222)>								
Risk genotypes								
0	reference		2,590	reference		984	**1.80 (1.04-3.11)**	**0.036327**
1	**8.76 (6.61-11.61)**	**< 2e-16**	2,541	**6.72 (4.66-9.70)**	**< 2e-16**	1,107	**22.05 (15.28-31.83)**	**< 2e-16**
p_{trend}					< 2e-16			
<Viscerally obese group, WHR > 0.85 (n=2,920)>								
Risk genotypes§								
0	reference		908	reference		479	0.96 (0.44-2.11)	0.928426
1	**6.02 (4.09 - 8.85)**	**< 2e-16**	977	**3.85 (2.39-6.20)**	**3.19e-08**	556	**10.82 (6.73-17.40)**	**< 2e-16**
p_{trend}					< 2e-16			
<Non-viscerally obese group, WST ≤ 88 cm (n=5,998)>								
Risk genotypes								
0	reference		2,206	reference		747	**2.02 (1.14-3.58)**	**0.01653**
1	**7.60 (5.65-10.21)**	**< 2e-16**	2,206	**5.82 (3.98-8.53)**	**< 2e-16**	839	**21.66 (14.76-31.77)**	**< 2e-16**
p_{trend}					< 2e-16			
<Viscerally obese group, WST > 88 cm (n=4,144)>								
Risk genotypes								
0	reference		1,353	reference		745	1.23 (0.65-2.32)	0.52740
1	**7.15 (5.15-9.93)**	**< 2e-16**	1,251	**5.06 (3.31-7.74)**	**6.47e-14**	795	**13.35 (8.77-20.32)**	**< 2e-16**
p_{trend}					< 2e-16			
<Active group, MET ≥ 10.0 (n=4,203)>								
Risk genotypes								
0	reference		1,481	reference		563	1.25 (0.62-2.55)	0.53390
1	**7.97 (5.64-11.28)**	**< 2e-16**	1,542	**5.85 (3.83-8.93)**	**3.13e-16**	617	**16.39 (10.65-25.22)**	**< 2e-16**
p_{trend}					< 2e-16			
<Inactive group, MET < 10.0 (n=5,939)>								
Risk genotypes								
0	reference		2,078	reference		929	**1.83 (1.07-3.13)**	**0.02792**
1	**6.97 (5.25-9.26)**	**< 2e-16**	1,915	**5.13 (3.50-7.52)**	**< 2e-16**	1,017	**17.67 (12.13-25.74)**	**< 2e-16**
p_{trend}					< 2e-16			
<Low fat-diet group, < 9.0% cal. from SFA (n=2,290)>								
Risk genotypes								
0	reference		788	reference		362	0.67 (0.22-2.06)	0.488644
1	**9.73 (5.96-15.90)**	**< 2e-16**	775	**4.50 (2.52-8.05)**	**3.99e-07**	365	**19.59 (11.09-34.60)**	**< 2e-16**
p_{trend}					< 2e-16			
<High fat-diet group, ≥ 9.0% cal. from SFA (n=7,852)>								
Risk genotypes								
0	reference		2,771	reference		1,130	**1.90 (1.19-3.04)**	**0.00715**
1	**6.92 (5.42-8.85)**	**< 2e-16**	2,682	**5.80 (4.19-8.03)**	**< 2e-16**	1,269	**16.50 (11.90-22.89)**	**< 2e-16**
p_{trend}					< 2e-16			

CI, confidence interval; CRC, colorectal cancer; HR, hazard ratio; MET, metabolic equivalent; OC, oral contraceptive; SFA, saturated fatty acids; WHR, waist-to-hip ratio; WST, waist circumference. Numbers in bold face are statistically significant. £The number of risk genotypes was defined on the basis of Kaplan-Meier's analysis as follows: 0 (none or 1 risk allele) and 1 (2 risk alleles). †Multivariate regression was adjusted by age at enrollment, education, body mass index, hip and waist circumferences (both in WHR strata; hip only in WST strata), waist-to-hip ratio (in MET and SFA strata), age at menopause, total months of breastfeeding, OC use (in total analysis), and exogenous estrogen plus progestin. **p* values were adjusted to correct for multiple testing via the Benjamini-Hochberg approach. §In the viscerally obese group (WHR), only the ONECUT2 rs4092465 GA genotype was analyzed.