**Supplementary materials**

## Title: Genetically predicted circulating C-reactive protein concentration and colorectal cancer survival: A Mendelian randomization consortium study

Xinwei Hua1,2, James Dai 1,2, Sara Lindström 1,2, Tabitha A Harrison 1, Yi Lin 1, Steven R Alberts3, Elizabeth Alwers 5, Sonja I Berndt 4, Hermann Brenner 5,6,7, Daniel D Buchanan 8,9,10, Peter T Campbell 11, Graham Casey 12, Jenny Chang-Claude 13,14, Steven Gallinger 15, Graham G Giles 16,17,18, Richard M Goldberg 19, Marc J Gunter 20, Michael Hoffmeister 5, Mark A Jenkins17, Amit Joshi 21,22, Wenjie Ma 22,23, Roger L Milne 16,17,18, Neil Murphy 20, Rish K Pai 24, Lori C Sakoda 1,25, Robert E Schoen 26, Qian Shi 27, Martha L Slattery 28, Mingyang Song22,23,29,31, Emily White1,2, Loic Le Marchand, 33 Andrew T Chan 30,31,32, Ulrike Peters 1,2, Polly A Newcomb1,2

**Supplementary methods**

**Estimation of European ancestry**

We first used Plink (v1.9) to conduct principal components analysis (PCA) to investigate population structure. Due to the low numbers of participants of other ancestries, we then restricted our analytic sample to participants with estimated European ancestry based on the PCA. The first two eigenvectors discriminated individuals based on self-identified race (**Supplementary Figure 1A**) and were used to select individuals with likely European ancestry. Participants with a value within one standard deviation of the median for the first and second eigenvectors were categorized as European ancestry (**Supplementary Figure 1B**) and retained for subsequent analyses. Within the subset of individuals with European ancestry, we recalculated the principal components. The first nine eigenvectors explained 73% of the genetic variation and were subsequently used as covariates in the analysis.

**Supplementary Table 1. Study-specific characteristics**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study Abbrev** | **Study Name** | **Study Design** | **Location** | **Genotyping platform** | **N Cases** | **Any deaths\* N (%)** | **CRC deaths\* N (%)** | **Age at diagnosis  Mean (SD)** | **Female (%)** | **Follow-up years, median (IQR)** | **Survival Ascertainment Method\*\*** |
| CCFR | Colon Cancer Family Registry | Case-control | USA, Australia, Canada | Illumina Human 1M or Human1M-Duo, Illumina Omni-1Quad, Illumina OncoArray, Custom Affymextrix Axiom array (1.3M SNPs) | 2453 | 821 (33.5) | 518 (21.1) | 58.7 (10.5) | 1154 (47) | 12.9  (6.1 - 15.7) | Registry linkage |
| CPSII | Cancer Prevention Study - II | Cohort | USA | Illumina OncoArray + Custom iSelect array, Custom Affymetrix Axiom Array (1.3M SNPs) | 816 | 312 (38.2) | 184 (22.5) | 76.4 (5.9) | 418 (51.2) | 5.2  (2.5 - 8.2) | Registry linkage |
| DACHS | Darmkrebs: Chancen der Verhutung durch Screening Study | Case-control | Germany | Illumina HumanCytoSNP, Illumina HumanOmniExpress, Illumina OncoArray | 2659 | 725 (27.3) | 537 (20.2) | 68.5 (10.6) | 1080 (40.6) | 3.3  (2.4 - 5) | Registry linkage |
| DALS | Diet Activity and Lifestyle Study | Case-control | USA | Illumina HumanCytoSNP, Illumina 550K, Illumina 610K | 1098 | 351 (32) | 210 (19.1) | 65.1 (9.8) | 492 (44.8) | 4.9  (3.5 - 6.6) | Registry linkage |
| EDRN | Early Detection Research Network | Case-cohort | USA | Illumina OncoArray + Custom iSelect Array | 190 | 19 (10) | 14 (7.4) | 61.6 (12) | 78 (41.1) | 3.4  (2.3 - 5.4) | Registry linkage |
| EPIC | European Prospective Investigation into Cancer | Cohort | Sweden | Illumina HumanOmniExpress + ExomeChip | 1821 | 588 (32.3) | 469 (25.8) | 63.4 (8.3) | 991 (54.4) | 3.4  (1.4 - 6.5) | Registry linkage |
| HPFS | Health Professionals Follow-up Study | Cohort | USA | Illumina HumanOmniExpress, Illumina HumanOmniExpress + ExomeChip | 344 | 151 (43.9) | 79 (23) | 71.8 (8.7) | 0 | 7.1  (3.4 - 12.4) | Active f/up |
| MCCS | Melbourne Collaborative Cohort Study | Cohort | Australia | Illumina OncoArray, Custom Affymetrix Axiom Array (1.3M SNPs) | 751 | 151 (20.1) | 77 (10.3) | 70.2 (9) | 359 (47.8) | 11.4  (6.6 - 15.5) | Registry linkage |
| N9741 | N9741 | Clinical trial | USA | Illumina HumanOmniExpress + ExomeChip | 426 | 405 (95.1) | 366 (85.9) | 60.7 (11.2) | 177 (41.5) | 1.6  (0.9 - 2.6) | Active f/up |
| NHS | Nurses' Health Study | Cohort | USA | Illumina HumanOmniExpress, Illumina HumanOmniExpress + ExomeChip | 587 | 199 (33.9) | 153 (26.1) | 69.1 (8.6) | 587 (100) | 7.6  (3.1 - 12.8) | Active f/up |
| PHS | Physician's Health Study | Cohort | USA | Illumina HumanOmniExpress | 323 | 170 (52.6) | 125 (38.7) | 70.6 (9.6) | 0 | 5.7  (2.1 - 13.1) | Active f/up |
| PLCO | Prostate, Lung, COlorectal, and Ovarian Cancer Screening Trial | Case-control | USA | Illumina HumanCytoSNP, Illumina HumanHap300 and HumanHap240S, Illumina 610K | 972 | 260 (26.7) | 174 (17.9) | 69.6 (6.3) | 416 (42.8) | 5.3  (2.7 - 8.5) | Active f/up |
| UKB | UK Biobank | Cohort | UK | Custom Affymetrix Axiom Array (1.3M SNPs) | 2877 | 731 (25.4) | 539 (18.7) | 64.5 (6.4) | 1204 (41.8) | 3.1  (1.8 - 4.9) | Registry linkage |
| VITAL | VITamins and Lifestyle Study | Cohort | USA | Illumina HumanCytoSNP | 270 | 109 (40.4) | 67 (24.8) | 69.8 (6.5) | 124 (45.9) | 4.9  (2.5 - 7.3) | Registry linkage |
| WHI | Women's Health Initiative | Cohort | USA | Illumina HumanCytoSNP, Illumina 550K, Illumina 610K | 1331 | 403 (30.3) | 296 (22.2) | 71.6 (7.2) | 1331 (100) | 3.8  (1.5 - 7.1) | Active f/up |
| \*All death events were censored at 10 years since diagnosis \*\*Registry linkage involves National Death Index, state cancer registries, state death records, or population registers with cause of death verified by death certificates; Active f/up involves death certificate and/or medical record review | | | | | | | | | | | | |

**Supplementary Table 2. Association between genetically determined CRP concentrations and CRC-specific survival, with and without adjustment of body mass index**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **52-SNP GRS** | **HD (95% CI) \*** | **P value** | **HD (95% CI) \*\*** | **P value** | | |
| Continuous | -1.12 (-2.65, 0.41) | 0.17 | -1.15(-2.76, 0.47) | 0.16 | | |
| By quartiles |  |  |  |  | | |
| Q1 (2.05,3.06] | 1.00 (Ref) |  | 1.00 (Ref) |  | | |
| Q2 (3.06,3.24] | 0.31 (-0.83, 1.46) | 0.60 | 0.31(-0.87, 1.49) | 0.61 | | |
| Q3 (3.24,3.41] | -0.54 (-1.67, 0.59) | 0.37 | -0.52(-1.71, 0.68) | 0.40 | | |
| Q4 (3.41,4.08] | -0.7 (-1.82, 0.41) | 0.23 | -0.73(-1.87, 0.41) | 0.21 | | |
| Abbreviation: CRC: colorectal cancer; CRP: C-reactive protein; GRS: genetic risk score; HD: Hazards difference per 100,000 person-year; CI: confidence interval  \*Adjust for age at diagnosis, sex, genotyping platform, study and principal components \*\*Adjust for the same covariate + body mass index | | | | |
|  | | | | | |

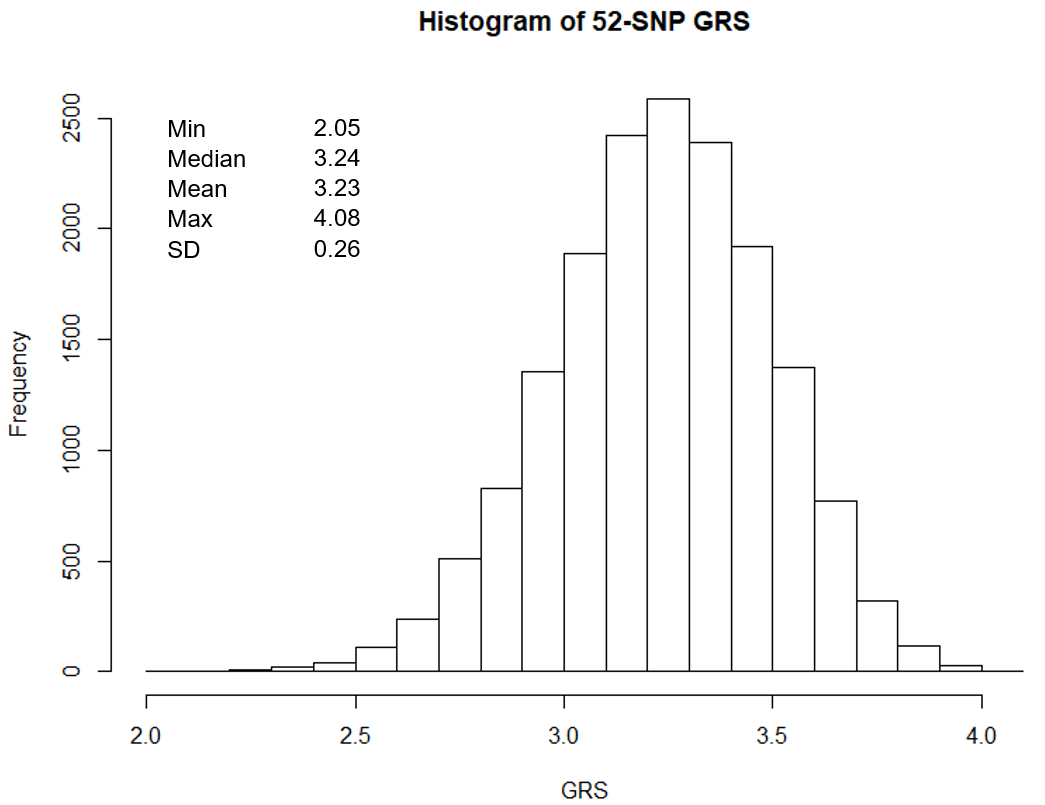
**Supplementary Figure 1.** Principal components analysis (PCA) used to define European genetic ancestry. (A) PCA plot colored by self-identified race. Values within one standard deviation of the median for the first and second eigenvectors was used to define individuals with European genetic ancestry. (B) individuals defined as European genetic ancestry are highlighted in blue. All others were excluded from further analysis.



A

B

**Supplementary Figure 2. Distribution of the 52-SNP genetic risk score for CRP**



**Supplementary program**

**R code for power calculation:**

library(survival)

library(timereg)

n <- 16918 ## sample size ##

out <- matrix(0,500,2)

for (i in 1:500) {

cat(i,"..")

z <- rnorm(n,0,1) ## this is GRS ##

x <- rnorm(n,sqrt(0.059)\*z,sqrt(0.941)) ## CRP, 5.9% variance explained by GRS, variance = 1- 0.059

#summary(lm(x~z))

#anova(lm(x~z))

#baseline hazard =3808/(16918\*10) =0.023

lamb <- 0.023+x\*(0.023\*0.25) ### this is to create effect size in additive scale; HD= 0.0056 ###

lamb <- ifelse(lamb<=0,0.001,lamb)

Tstart <- runif(n,0,1) ### stagger entry/enrollment during year 1 ##

ftime <- rexp(n,lamb)

ctime <- Tstart+ftime

tcut <- 10 ## follow up 10 years ##

ftime <- ifelse(ctime<tcut,ctime,tcut-Tstart)

censor<- ifelse(ctime<tcut,1,0)

out[i,2] <- mean(censor) ## event rate in 10 years is around 0.23

fit1 <- aalen(Surv(ftime,censor)~z)

out[i,1] <- fit1$pval.testBeq0[2]

print(out[i,])

}

mean(out[,1]<0.05) ## power

mean(out[,2]) ## event rate on average