Mutagenesis, 2025, **40**, 48–60
https://doi.org/10.1093/mutage/geae008
Advance access publication 5 March 2024
Original Manuscript

OXFORD

Editor's choice

Original Manuscript

Inflammation and gut barrier function-related genes and colorectal cancer risk in western European populations

Hannah B. Mandle[1], Mazda Jenab[2], Marc J. Gunter[2,3], Anne Tjønneland[4,5], Anja Olsen[5,6], Christina C. Dahm[6], Jie Zhang[6], Pierre-Emmanuel Sugier[7,8], Joseph Rothwell[7], Gianluca Severi[7], Rudolf Kaaks[9], Verena A. Katzke[9], Matthias B. Schulze[10,11], Giovanna Masala[12], Sabina Sieri[13], Salvatore Panico[14], Carlotta Sacerdote[15], Catalina Bonet[16,17], Maria-Jose Sánchez[18,19,20,21], Pilar Amiano[20,22], José María Huerta[20,23], Marcela Guevara[20,24,25], Richard Palmqvist[26], Thyra Löwenmark[26], Aurora Perez-Cornago[27], Elisabete Weiderpass[28], Alicia K. Heath[3], Amanda J. Cross[3], Paolo Vineis[29,30], David J. Hughes[31,‡,*], Veronika Fedirko[1,32,‡,*]

[1]Department of Epidemiology, Emory Rollins School of Public Health, Atlanta, GA 30322, USA
[2]Nutrition and Metabolism Branch, International Agency for Research on Cancer (IARC-WHO), 69372 Lyon, France
[3]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, SW7 2AZ, UK
[4]Diet, Cancer and Health, Danish Cancer Society Research Center, 2100 Copenhagen, Denmark
[5]Department of Public Health, University of Copenhagen, 1353 Copenhagen K, Denmark
[6]Department of Public Health, Aarhus University, DK-8000 Aarhus C, Denmark
[7]Université Paris-Saclay, UVSQ, Inserm 'Exposome and Heredity' team, CESP U1018, 94807 Villejuif Cedex, France
[8]Laboratoire de Mathématiques et de leurs Applications de Pau E2S UPPA, CNRS, 64013 Pau Cedex, France
[9]Division of Cancer Epidemiology, German Cancer Research Center, DKFZ, 69120 Heidelberg, Germany
[10]Department of Molecular Epidemiology, German Institute of Human Nutrition, Potsdam-Rehbruecke, 14469 Nuthetal, Germany
[11]Institute of Nutritional Science, University of Potsdam, 14469 Nuthetal, Germany
[12]Institute for Cancer Research, Prevention and Clinical Network (ISPRO), 50139 Florence, Italy
[13]Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, 20133 Milan, Italy
[14]Dipartimento Di Medicina Clinica E Chirurgia, Federico II University, 80131 Naples, Italy
[15]Unit of Cancer Epidemiology, AOU Città della Salute e della Scienza University Hospital, 10126 Turin, Italy
[16]Unit of Nutrition and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology (ICO), L'Hospitalet de Llobregat, 0890x Barcelona, Spain
[17]Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, 0890x Barcelona, Spain
[18]Escuela Andaluza de Salud Pública (EASP), 18011 Granada, Spain
[19]Instituto de Investigación Biosanitaria ibs. 18011 Granada, Spain
[20]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), 28029 Madrid, Spain
[21]Department of Preventive Medicine and Public Health, University of Granada, 18071 Granada, Spain
[22]Ministry of Health of the Basque Government, Sub Directorate for Public Health and Addictions of Gipuzkoa, BioGipuzkoa Health Research Institute, Epidemiology of Chronic and Communicable Diseases Group, 20014 Donostia – San Sebastian, Spain
[23]Department of Epidemiology, Murcia Regional Health Council-IMIB, 30120, El Palmar, Murcia, Spain
[24]Instituto de Salud Pública y Laboral de Navarra, 31003 Pamplona, Navarra, Spain
[25]Navarra Institute for Health Research (IdiSNA), 31008 Pamplona, Navarra, Spain
[26]Department of Medical Biosciences, Umea University, 901 87 Umeå, Sweden
[27]Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK
[28]Office of the Director, International Agency for Research on Cancer, 69366 Lyon Cedex 07, France
[29]MRC Centre for Environment and Health, School of public Health, Imperial College London, London W2 1PG, UK
[30]Italian Institute for Genomic Medicine (IIGM), 10060 Candiolo TO, Italy
[31]Cancer Biology and Therapeutics Group, School of Biomolecular and Biomedical Science, UCD Conway Institute, University College Dublin, Dublin 4, Ireland
[32]Department of Epidemiology, University of Texas M. D. Anderson Cancer Center, 77030 Houston, TX, USA

‡These authors contributed equally to this work and share senior authorship.
*Corresponding author. M. D. Anderson Cancer Center 1515 Holcombe Blvd.Houston, Texas 77030, USA. E-mail: vfedirko@mdanderson.org (V.F.); University College Dublin, School of Biomolecular and Biomedical Science, Conway Institute Belfield Dublin 4, Ireland. E-mail: david.hughes@ucd.ie (D.J.H)

Received 22 September 2023; accepted 4 April 2024

Abstract

Gut barrier dysfunction and related inflammation are known to be associated with the development and progression of colorectal cancer (CRC). We investigated associations of 292 single-nucleotide polymorphisms (SNPs) from 27 genes related to endotoxins/lipopolysaccharide (LPS) sensing and tolerance, mucin synthesis, inflammation, and Crohn's disease with colon and rectal cancer risks. Incident CRC cases (N = 1374; colon = 871, rectum = 503) were matched 1:1 to controls nested within the European Prospective Investigation into Cancer and Nutrition cohort. Previously measured serum concentrations of gut barrier function and inflammation biomarkers (flagellin/LPS-specific immunoglobulins and C-reactive protein [CRP]) were available for a sub-set of participants (N_{cases} = 1001; $N_{controls}$ = 667). Forty-two unique SNPs from 19 different genes were associated with serum biomarkers at $P_{unadjusted} \leq 0.05$ among controls. Among SNPs associated with a gut permeability score, 24 SNPs were in genes related to LPS sensing and mucin synthesis. Nine out of 12 SNPs associated with CRP were in genes related to inflammation or Crohn's disease. *TLR4* was associated with colon cancer at the SNP level (nine SNPs, all $P_{unadjusted} \leq 0.04$) and at the gene level ($P_{unadjusted} \leq 0.01$). *TLR4* rs10759934 was associated with rectal cancer but not colon cancer. Similarly, *IL10* was associated with rectal cancer risk at an SNP and gene level (both $P_{unadjusted} \leq 0.01$), but not colon cancer. Genes and SNPs were selected *a priori*; therefore, we present unadjusted *P*-values. However, no association was statistically significant after multiple testing correction. This large and comprehensive study has identified gut barrier function and inflammation-related genes possibly contributing to CRC risk in European populations and is consistent with potential etiological links between host genetic background, gut barrier permeability, microbial endotoxemia, and CRC development.

Keywords: single nucleotide polymorphism; gut barrier; inflammation; colorectal neoplasms; incidence

Introduction

The human gut microbiome, within the colon segments and rectum, contains roughly 3000 different bacterial species as identified and isolated in human feces [1]. The interaction between the gut microbiome and the gastrointestinal tract, an estimated total area of 400 m^2, constitutes the body's greatest exposure to the external environment [2]. The first line of defense against endotoxins and foreign antigens in the gut is comprised of a thick inner mucus layer followed by the epithelium. Maintenance of bacterial eubiosis as well as integrity of the gut barrier may play a crucial role in the prevention of several diseases impacted by these factors such as colon and rectal cancers (CRCs). Pathogenic bacteria (and some commensals such as *Bacteroides fragilis*) have been observed to penetrate the mucus layer via degradation of glycoproteins [3], evolutionary advances in motility [4,5], or interfering with commensal triggers of mucus production [6]. Increased permeability of the intestinal mucosal barrier may lead to the translocation of viable bacteria and microbial products to the lamina propria and systemic bloodstream [6,7]. Exposure to pathogenic bacteria and pathogen-associated molecular patterns (PAMPs) can result in metabolic endotoxemia and systemic inflammation [8,9], which, in turn, can promote tumorigenesis via upregulation of cell proliferation, resistance to apoptosis, increased angiogenesis, and other mechanisms [10]. Chronic exposure to pathogens and subsequent inflammation can further damage the function and integrity of the barrier [11].

The immunological barrier within the epithelium and lamina propria follows the physical barrier. In this layer, toll-like receptors (TLRs) play a key role by recognizing lipopolysaccharides (LPS), a PAMP, and an integral part of the outer membrane of gram-negative bacteria cell walls [12]. LPS-sensing by LPS binding protein (LBP) and CD14 in the mucosal layer activates TLR4 which induces downstream nuclear factor κβ (NF-κB)-mediated production and secretion of pro-inflammatory cytokines [13]. Activation of the NF-κB signaling pathway was found to contribute to colon cancer development and progression via transcriptional upregulation of cell proliferation and angiogenesis, inhibition of apoptosis, and overexpression of cyclooxygenase-2, which further promotes inflammation and cell proliferation [13–16].

Composition and function of the gut barrier and the degree of immune response to PAMPs may be modulated by variations in a broad array of genes related to LPS/endotoxin recognition, mucin synthesis, and inflammation. Despite experimental and observational evidence of the importance of gut barrier integrity in association with CRC and the polymorphic nature of these genes, few studies have investigated these functional genes and respective signaling pathways in relation to CRC risk. Of the available studies, most had small sample sizes (<200 cases) in specific populations with a focus on minimal genes and few single nucleotide polymorphisms (SNPs) resulting in inconsistent findings [17–20].

The etiological roles the gut barrier and immune response play in CRC development may also vary by tumor molecular subtype, primary tumor site, and sex. Lifestyle factors that can alter gut microbiota and barrier function, such as physical activity, diet, smoking, and body mass index (BMI), have been found to be strong risk factors for colon cancer while a healthier lifestyle has been found to have less of an impact in preventing rectal cancer [21,22]. In addition, tumor site prevalence varies by sex with proximal colon tumors being more common in women and rectal tumors occurring more frequently in men [23].

To date, no studies have comprehensively investigated the individual and collective associations of gut-barrier and inflammatory-related genes with CRC risk alone and in combination with circulating biomarkers of systemic exposure to PAMPs. Therefore, we investigated whether common genetic variation in genes related to LPS/endotoxin recognition, mucin synthesis, and inflammation are related to circulating overall gut permeability, LPS, flagellin, and C-reactive protein (CRP) levels. We also investigated whether genetic variations at the SNP, gene and pathway level, alone and in combination with these circulating biomarker levels, are associated with CRC risk and whether these associations differ by tumor site and/or sex in a large Western European prospective cohort study.

Materials and methods

Study population

We used a case-control design nested within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, a large prospective study with over 520,000 men and women aged 35–70 years enrolled from 23 centers in 10 Western European countries (Denmark, France, Greece, Germany, Italy, the Netherlands, Norway, Spain, Sweden, and UK). The methods of the EPIC study have been described in detail elsewhere [24,25]. Individuals who were eligible for the study were from the general population of a specific geographical area, town, or province. Exceptions included the French sub-cohort, which is based on members of the

health insurance system or state-school employees, and the Utrecht (Netherlands) sub-cohort, which is based on women who underwent screening for breast cancer. Between 1992 and 2000, standardized lifestyle and personal history information, anthropometrics, and blood samples were collected from most participants at recruitment. Diet over the previous 1 year was measured at baseline by validated country-specific dietary questionnaires developed to ensure high compliance and better measures of local dietary habits [24]. Blood samples were stored at the International Agency for Research on Cancer (IARC, Lyon, France; –196°C, in liquid nitrogen) for all countries except Denmark (–150°C, in nitrogen vapor) and Sweden (in –80°C freezers). The present analysis is based on participant data from all centers except for Norway (blood samples were only recently collected; few CRCs diagnosed after blood sampling), the Malmö center of Sweden (no available serum samples), and Greece (excluded from the current analysis due to data restriction issues). Written informed consent was provided by all study participants. The EPIC study was approved by the Ethical Review Board of the IARC and the Institutional Review Board of each participating EPIC center. Written consent was obtained from all EPIC participants at enrollment into the study.

Cancer incidence and vital status follow-up

Cancer incidence was determined through record linkages with regional cancer registries (Denmark/Italy/the Netherlands/Norway/Spain/Sweden/UK; complete up to December 2006) or via a combination of methods, including the use of health insurance records, contacts with cancer and pathology registries, and active follow-up through study subjects and their next-of-kin (France/Germany/Naples/Greece; complete up to June 2010).

Vital status follow-up (98.5% complete) was collected by record linkage with regional and/or national mortality registries in all countries except France and Germany, where data are collected through an active follow-up. Censoring dates for complete follow-up were between June 2005 and June 2009 in Denmark, the Netherlands, Spain, the UK, Sweden, Norway, and Italy. In Germany, Greece, and France follow-up was based on a combination of methods, including health insurance records, cancer and pathology registries, and active follow-up through study subjects and their next-of-kin. In these centers, the end of follow-up was defined as the last known date of contact, or the date of death, whichever came first. The last update of endpoint information occurred between December 2007 and December 2009.

Nested case-control design and participant selection

Case ascertainment and selection

Colorectal cancer (CRC) cases were selected among participants who developed colon (C18.0–C18.7, according to the ICD-10), rectum (C19–C20), and overlapping/unspecified origin tumors (C18.8 and C18.9). Cancers of the anus were excluded. CRC is defined as the combination of colon and rectal cancers (C18–C20).

A total of 1374 first-incident CRC cases (colon cancer = 871; rectal cancer = 503) were identified. The number of cases for analyses involving gut barrier biomarker serum levels was 1001 because of missing or unobtainable, previously collected measurements [26] (France = 5, Italy = 78, Spain = 44, UK = 54, The Netherlands = 29, Germany = 66, Sweden = 88, Denmark = 9). The number of cases for analyses involving inflammation serum levels was 1082 (missing: France = 7, Italy = 70, Spain = 40, UK = 43, The Netherlands = 21, Germany = 47, Sweden = 36, Denmark = 28). Nine-hundred and forty-six (68.9%) cases had both gut barrier and inflammation serum measurements.

Control selection

Controls were selected (1:1) by incidence density sampling from all cohort members alive and not having a reported cancer at the time of diagnosis of the cases and were matched by age (±6 months at recruitment), sex, study center, time of the day at blood collection, and fasting status at the time of blood collection (less than 3 h, 3–6 h, and more than 6 h). Women were further matched by menopausal status (pre-/post-/perimenopausal, and unknown) and for pre-menopausal women phase of menstrual cycle at time of blood collection, and usage of postmenopausal hormone therapy at time of blood collection (yes/no, regardless of menopausal status). The additional matching criteria for women were required for other studies that were being done using the same matched case-control sets. Three control samples failed the genotyping and were not included in the analysis, resulting in 1371 controls. The number of controls for analyses involving gut barrier biomarker serum levels was 667 because of missing or unobtainable, previously collected measurements (France = 14, Italy = 73, Spain = 44, UK = 67, The Netherlands = 31, Germany = 75, Sweden = 87, and Denmark = 313) [26]. The number of controls for analyses involving inflammation was 688 (missing: France = 19, Italy = 74, Spain = 43, UK = 67, The Netherlands = 47, Germany = 69, Sweden = 45, and Denmark = 319). Five-hundred and eighty-eight (42.9%) controls had both gut barrier and inflammation serum measurements.

Serum biomarker assessment

Serum concentration measurements had been conducted previously on the cases and controls [27–30]. Consistent with previous literature [29], to assess systemic exposure to PAMPs, proxy measures for gut barrier function and permeability, specific immunoglobulins against LPS and flagellin were measured. Serum LPS- and flagellin-specific immunoglobulin G (IgG) and immunoglobulin A (IgA) levels were quantified by ELISA at Georgia State University. All intra-assay coefficients of variation were between 3.8% and 6.8%. An overall gut permeability score was calculated as a sum of LPS- and flagellin-specific immunoglobulin levels (LPS IgG + LPS IgA + flagellin IgG + flagellin IgA), total LPS and flagellin were calculated as sums of the LPS- or flagellin-specific immunoglobulin levels only. To assess systemic inflammation, CRP concentrations were quantified using a high-sensitivity assay (Beckman-Coulter). The intra-assay coefficients of variation were 6.0%–6.5% at various concentrations of CRP.

SNP selection, genotyping, and quality control

Genomic DNA was extracted from whole blood samples using conventional methods. We used the custom GoldenGate Universal-32 3,072-plex assay kit (Illumina, Ca, USA) to genotype 397 genetic variants within genes known and proposed to be involved with (i) endotoxin and LPS sensing (*TLR4, TNFRSF1B, LBP, LPS tolerance, LY96, CD14, MYD88*), (ii)

mucin encoding (*ABCB1, CAMP/ZNF589, CD38, MUC1, MUC12, MUC17, MUC6/MUC2*), and (iii) inflammation (*ALOX5, IL10, IL10R, IL12A, IL12B, IL2/IL21, IL6, IFNG, TNF, IL1A/IL1B, IL8, IL18, NFKB1, RELA (p65)*) [please see Supplementary Table 1 for a complete list of genes and SNPs]. In addition, from previously published GWAS studies [31,32], we a priori selected SNPs related to Crohn's disease and their potential response to vitamin D, which has been associated with CRC risk [33]. The custom GoldenGate assay was designed using the Illumina online Assay Design Tool in May 2012. SNP genotype data-set for CEU population (Utah residents with Northern and Western European ancestry; HapMap Data Rel 28 Phase II + III, August 10, on NCBI B36 assembly, dbSNP b126) were loaded in the Haploview program (Broad Institute, MIT and Harvard, Cambridge, MA, USA) and SNPs with minor allele frequencies (MAFs) greater than 5% and the r^2 linkage disequilibrium (LD) statistic of 0.8 were selected as tagging SNPs (tagSNPs). Additionally, we searched published literature for previously reported functional and regulatory SNPs in the genes of interest and included them in genotyping irrespective of MAFs or r^2 with other SNPs. Genotyping was performed by the Genetics Laboratory at Imperial College London. After excluding 105 SNPs [65 (16.4%) that failed genotyping, 11 (2.8%) that failed to satisfy the Hardy–Weinberg criterion (Supplementary Table 1), 25 (6.3%) missing in more than 20% of genotyped samples, and 4 (1.0%) that were monomorphic], 292 SNPs were included in the analysis. All genotyping underwent standard quality control including concordance checks for blinded duplicates and examination of sample and SNP call rates. The lowest reproducibility frequency across 62 replicate samples was 0.98. The call rate was 95% for all samples and 95% for all SNPs.

Statistical analysis

The associations between serum biomarkers and genetic variants (coded as 0, 1, 2 corresponding to the number of minor alleles) were assessed among controls using linear regression models adjusted for age, sex, and center. Further adjustments for BMI, smoking status, and physical activity did not change the results substantially and, therefore, were removed from the final mode. We used unconditional logistic regression analysis to assess the association of individual SNPs with colon and rectal cancer risks, adjusting for age (continuous), sex, and study center. Results were similar when we used conditional logistic regression on 1325 complete case-matched sets. Heterogeneity among statistically significant estimates for colon and/or rectal cancer risks was assessed using a meta-analytic approach and I^2 statistic. We assumed an additive genetic model but also tested dominant and recessive models as the underlying genetic model for these SNPs is unknown. Adjustment for BMI (continuous), smoking status (never, former, current smokers, missing), physical activity (active, moderately active, moderately inactive, and inactive), alcohol intake (continuous), hormone therapy, and menopausal status did not substantially change the results, so we did not include them the final statistical model. To conduct sub-analyses assessing interaction by sex, we used an additive model and a Wald χ^2 test statistic.

To examine the associations between genes (a combination of SNPs) and genetic pathways (a combination of genes) and CRC risk, we used the Adaptive Rank Truncated Product (ARTP) method [17] as implemented in the first step (no interaction) of the R package PIGE (http://cran.r-project.org/web/packages/PIGE/index.html). This method can combine associations of SNPs in each gene (or from the genes in a pathway) to provide a *P* value at the gene or pathway level, respectively. Genetic markers in high LD ($r^2 \geq 0.8$) were excluded using the AdaJoint R package (https://cran.r-project.org/web/packages/ARTP2). Analyses were conducted on all participants as well as among only the participants who had serum measurements. The results did not differ between participants with gut barrier serum measurements and participants with inflammation serum measurements, therefore we chose to report the results for only the gut barrier group. To investigate the multiplicative interaction between the genes and genetic pathways with gut permeability and inflammation biomarkers on CRC risk, we used the modified ARTP method as implemented in the R package PIGE. We further assessed the association of biomarkers (per tertile) with CRC risk by genotype.

All statistical tests were two-sided with *P*-values < 0.05 considered statistically significant (SAS software, version 9.2; SAS Institute, Cary/NC; R, R Foundation for Statistical Computing, Vienna/Austria). All *P*-values at the SNP, gene, or pathway levels were corrected for multiple testing for the number of statistical tests being performed at each level, respectively, by using the false discovery rate (Benjamini–Hochberg or BH) method [34].

Results

Baseline characteristics of cases and controls

Selected baseline characteristics of the CRC cases and matched controls are shown in Table 1. The mean age at blood sample collection of cases and controls was 58.5 years (SD: 7.2). On average, CRC cases had 4 years between blood collection and the time of diagnosis. The dataset included 871 colon cancer cases and 503 rectal cancer cases. Gut barrier function and inflammation serum biomarkers did not meaningfully differ between cases and controls.

SNPs, genes, and pathways

We examined associations of SNPs in genes within pathways relating to endotoxin and LPS sensing (genes = 7, SNPs = 78), mucin encoding (genes = 7, SNPs = 66), inflammation (genes = 14, SNPs = 133), and 15 SNPs in the intergenic regions previously found to be associated with Crohn's disease [31] with circulating serum biomarker levels and colon and rectal cancer risks.

Associations with circulating serum biomarkers

In total, 54 total (42 unique) SNPs in 19 genes and regions related to LPS/endotoxins, mucins, inflammation, or Crohn's disease were associated with serum biomarkers of gut barrier function and/or inflammation at an unadjusted $P \leq 0.05$ among controls (Table 2). For the gut barrier permeability markers (LPS and flagellin), 24 (57%) of the 42 associated SNPs were in genes related to gut barrier function (LPS/endotoxin sensing and mucin encoding). Nine (75%) of the 12 SNPs associated with CRP were in genes related to inflammation or associated with Crohn's disease. None of these SNPs were statistically significantly associated with gut permeability or inflammation after BH correction. The

Table 1. Selected baseline characteristics of incident CRC cases (overall and by tumor site) and their matched controls, the EPIC study, 1992–2003.

Baseline characteristic	Controls	CRC cases	Colon cases	Rectal cases
	N = 1371	*N* = 1374	*N* = 871	*N* = 503
Women, *n* (%)	683 (49.82)	687 (50.00)	464 (53.27)	223 (44.33)
Mean age at blood collection, yrs, SD	58.50, 7.20	58.45, 7.20	58.75, 7.40	57.90, 6.80
Mean years of follow-up, yrs, SD		4.18, 2.29	4.16, 2.31	4.23, 2.25
Smoking status, *n* (%)[a]				
Never	575 (41.94)	560 (40.76)	372 (42.71)	188 (37.38)
Former	445 (32.46)	462 (33.62)	291 (33.41)	171 (34.00)
Current	336 (24.51)	340 (24.75)	200 (22.96)	140 (27.83)
Physical activity, *n* (%)[a]				
Inactive	181 (13.20)	200 (14.56)	127 (14.58)	73 (14.51)
Moderately inactive	353 (25.75)	383 (27.87)	247 (28.36)	136 (27.04)
Moderately active	584 (42.60)	565 (41.12)	362 (41.56)	203 (40.36)
Active	143 (10.43)	123 (8.95)	70 (8.04)	53 (10.54)
BMI, kg/m^2, SD	26.24 (3.79)	26.73, 4.22	26.82, 4.38	26.55, 3.92
Gut barrier biomarker serum measurements, *n* (%)	667 (48.65)	1001 (72.85)	629 (45.78)	372 (73.96)
Gut permeability score, OD, SD	6.82, 2.08	6.47, 2.04	6.60, 2.02	6.25, 2.04
Lipopolysaccharide, OD, SD	3.37, 1.16	3.19, 1.17	3.27, 1.18	3.06, 1.14
Flagellin, OD, SD	3.45, 1.25	3.28, 1.21	3.33, 1.20	3.19, 1.23
Inflammation biomarker serum measurements, *n* (%)	688 (50.18)	1082 (78.75)	686 (49.93)	396 (78.73)
C-reactive protein, mg/l, SD	3.74, 6.24	4.33, 6.16	4.71, 6.65	3.67, 5.15
Country, *n* (%)[a]				
Italy	196 (14.30)	202 (14.70)	144 (16.53)	58 (11.53)
France	29 (2.12)	28 (2.04)	22 (2.53)	6 (1.19)
Spain	142 (10.36)	146 (10.63)	101 (11.60)	45 (8.95)
UK	250 (18.23)	240 (17.47)	166 (19.06)	74 (14.71)
The Netherlands	158 (11.52)	153 (11.14)	99 (11.37)	54 (10.74)
Germany	169 (12.33)	179 (13.03)	110 (12.63)	69 (13.72)
Sweden	87 (6.35)	88 (6.40)	55 (6.31)	33 (6.56)
Denmark	340 (24.80)	338 (24.60)	174 (19.98)	164 (32.60)

Abbreviations: EPIC, European Prospective Investigation into Cancer and Nutrition; SD, standard deviation; CRC, colorectal cancer; BMI, body mass index; *N*, sample size; yrs, years; OD, optical density.

[a]Percent missing is not shown; therefore, the total percentages do not add up to 100%.

Table 2. SNPs in genes related to the gut barrier function were found to be associated (unadjusted *P*-value < 0.05) with gut permeability score, LPS, flagellin, and CRP among controls, the EPIC study, 1992–2003.

Biomarker	Gene	SNP	*N* controls	β (95% CI)[a]	*P*[b]
Gut permeability score	*ABCB1*	rs1055302	656	−0.33 (−0.65, −0.01)	*0.04*
	ALOX5	rs2291427	667	−0.25 (−0.49, −0.02)	*0.04*
	ALOX5	rs7896431	641	0.24 (0.02, 0.45)	*0.03*
	Crohn's Disease[c]	rs6908425	649	−0.30 (−0.57, -0.03)	*0.03*
	IL1A/IL1B	rs419598	665	−0.30 (−0.55, −0.05)	*0.02*
	MUC6/MUC2	rs10794288	664	−0.32 (−0.60, −0.05)	*0.02*
	MUC6/MUC2	rs10902088	662	−0.31 (−0.59, −0.04)	*0.03*
	MUC6/MUC2	rs41411848	663	−0.30 (−0.57, −0.02)	*0.03*
	MUC6/MUC2	rs57737240	665	−0.29 (−0.57, −0.02)	*0.04*
	MUC6/MUC2	rs7119740	657	−0.56 (−1.01, −0.11)	*0.02*
	NFKB1	rs11940017	665	0.65 (0.10, 1.20)	*0.02*
	TNF	rs909253	663	−0.25 (−0.48, −0.01)	*0.04*
LPS	*ABCB1*	rs1055302	656	−0.18 (−0.35, 0.00)	*0.04*
	ABCB1	rs1858923	660	0.13 (0.01, 0.25)	*0.04*
	ALOX5	rs10751382	657	−0.15 (−0.27, −0.02)	*0.03*
	ALOX5	rs2291427	667	−0.15 (−0.28, −0.02)	*0.02*
	ALOX5	rs3780914	644	−0.14 (−0.26, −0.03)	*0.02*

Table 2. Continued

Biomarker	Gene	SNP	*N* controls	*β* (95% CI)[a]	*P*[b]
	ALOX5	rs7896431	641	0.15 (0.03, 0.27)	*0.01*
	CD14	rs2563298	666	−0.18 (−0.32, −0.05)	*0.01*
	CD14	rs2569190	665	0.15 (0.03, 0.27)	*0.02*
	CD14	rs2569191	651	0.14 (0.02, 0.27)	*0.03*
	Crohn's Disease	rs6908425	649	−0.18 (−0.33, −0.03)	*0.02*
	IL6	rs1474347	658	−0.13 (−0.26, −0.01)	*0.04*
	IL6	rs1474348	657	−0.13 (−0.26, −0.01)	*0.04*
	MUC12	rs6980143	662	0.14 (0.01, 0.27)	*0.04*
	MUC17	rs10259584	639	0.25 (0.03,0.46)	*0.03*
	MUC6/MUC2	rs10794288	664	−0.16 (−0.31, −0.01)	*0.03*
	MUC6/MUC2	rs10902088	662	−0.18 (−0.33, −0.03)	*0.02*
	MUC6/MUC2	rs2071174	657	−0.16 (−0.29, −0.03)	*0.02*
	MUC6/MUC2	rs4077759	658	−0.18 (−0.31, −0.05)	*0.01*
	MUC6/MUC2	rs41411848	663	−0.16 (−0.31, −0.01)	*0.04*
	MUC6/MUC2	rs7119740	657	−0.26 (−0.51, −0.02)	*0.04*
	NFKB1	rs10489113	665	−0.17 (−0.32, −0.01)	*0.03*
	NFKB1	rs11940017	665	0.35 (0.05, 0.66)	*0.02*
	NFKB1	rs7665090	653	0.16 (0.04, 0.28)	*0.01*
	TNFRSF1B	rs4846100	665	−0.18 (−0.30, −0.06)	*<0.01*
Flagellin	*IL12B*	rs7709212	662	0.15 (0.01, 0.30)	*0.04*
	IL1A/IL1B	rs1143627	666	−0.14 (−0.28, −0.01)	*0.04*
	IL1A/IL1B	rs419598	665	−0.18 (−0.33, −0.03)	*0.02*
	IL8	rs4073	657	0.16 (0.03, 0.29)	*0.01*
	MUC6/MUC2	rs7119740	657	−0.29 (−0.56, −0.02)	*0.04*
	TNF	rs909253	663	−0.15 (−0.29, −0.01)	*0.04*
CRP	*ALOX5*	rs934187	684	0.11 (0.00, 0.22)	*0.04*
	CD38	rs3796878	680	0.19 (0.06, 0.33)	*0.01*
	Crohn's Disease	rs17582416	687	−0.12 (−0.23, −0.01)	*0.04*
	Crohn's Disease	rs4613763	686	0.19 (0.02,0.36)	*0.03*
	IFNG	rs1861493	677	−0.15 (−0.27, −0.03)	*0.01*
	IFNG	rs2069727	654	0.11 (0.00, 0.22)	*0.04*
	IFNG	rs971545	685	0.12 (0.01, 0.22)	*0.03*
	IL10	rs3024505	668	−0.22 (−0.37, −0.06)	*0.01*
	LBP	rs5741812	657	−0.19 (−0.37, 0.00)	*0.04*
	LPS tolerance	rs12652669	686	−0.31 (−0.55, −0.08)	*0.01*
	NFKB1	rs3774934	684	0.19 (0.01, 0.38)	*0.04*
	NFKB1	rs3821958	683	−0.13 (−0.24, −0.01)	*0.03*

Abbreviations: SNP, single nucleotide polymorphism; LPS, lipopolysaccharide; IgG, Immunoglobulin G; IgA, Immunoglobulin A; CRP, C-reactive protein; EPIC, European Prospective Investigation into Cancer and Nutrition, CI, confidence interval.

[a]Adjusted for age at blood collection, sex, and study center.

[b]Not statistically significant after BH method for multiple testing correction.

[c]SNPs previously found in GWAS to be associated with Crohn's disease [62].

associations for all SNPs with each biomarker are shown in Supplementary Table 2.

Associations with cancer risks

Colon cancer

In Table 3 the ORs for 12 statistically significant SNPs associated with colon cancer risk defined by $P_{unadjusted} < 0.05$ are shown. All nine SNPs related to endotoxins were in the *TLR4* gene. All SNP associations with colon cancer risk are presented in Supplementary Table 3; no association remained statistically significant after BH correction.

Associations with colon cancer risk at the gene and pathway level are presented in Table 4. At the whole gene variation level, and consistent with the individual SNP-level findings in Table 3, *TLR4* was statistically significantly associated with colon cancer risk ($P_{unadjusted} = 0.006$). However, the association between *TLR4* and colon cancer risk was not observed to be modified by any of the gut barrier or inflammation serum biomarkers. An interaction *P*-value of 0.021 was observed for *MUC6/MUC2* and flagellin; this was the only statistically significant finding in the mucin pathway. Within the inflammation pathway, *ALOX5* gene variation was associated with colon cancer risk when interacting with both overall circulating gut permeability and LPS alone (all $P_{unadjusted} \leq 0.01$). Similarly, *IL10R* had statistically significant interaction *P*-values for gut permeability and flagellin alone

(all $P_{unadjusted} = 0.03$). Crohn's disease was the only pathway statistically significantly associated with colon cancer risk ($P_{unadjusted} = 0.01$); however, this association was only observed when the pathway was considered without circulating biomarker interaction. No gene or pathway was statistically significantly associated with colon cancer risk after BH correction. All corrected *P*-values are shown in Supplementary Table 4.

Rectal cancer

Five statistically significant ($P_{unadjusted} < 0.05$) SNPs were associated with rectal cancer risk (Table 3). A SNP in the *TLR4* gene was also associated with rectal cancer but differed from those observed for colon cancer risk. One SNP in the inflammation pathway was associated with rectal cancer (*IL10*). *IL10* was also statistically significantly associated with rectal cancer risk when analyzed at the gene level and without any biomarker interaction ($P_{unadjusted} < 0.01$; Table 5). Additional associations within the inflammation pathway included *IL8* when interacting with flagellin ($P_{unadjusted} = 0.05$) and *NF-κB* when interacting with CRP ($P_{unadjusted} < 0.02$). The only statistically significant association with rectal cancer not within the inflammation pathway was an interaction of flagellin and *CD38* ($P_{unadjusted} < 0.05$). No gene associations were statistically significant after BH correction and no pathways were statistically significantly associated with rectal cancer risk, alone or with interaction of any biomarker, prior to *P*-value correction (Table 5 and Supplementary Table 4).

Exploratory Sub-analyses

We assessed SNP-level effect modification by sex and observed statistically significant sex-specific differences in colon and rectal cancer risks among genes in the LPS/endotoxin, mucin, and inflammation pathways.

For colon cancer, all SNPs with statistically significant interaction by sex within the *TNFRSF1B* and *LBP* genes were associated with a decrease in risk among men and an increase in risk among women per every one change in a minor allele. *LY96* was the only gene within the LPS/endotoxin pathway in which a SNP with statistically significant interaction by sex was associated with a decreased risk of colon cancer among women and an increased risk among men (Supplementary Table 5). Three SNPs in two genes were observed to have statistically significant interactions by sex in the mucin pathway (*ABCB1* and *MUC6/MUC2*) as well as in the inflammation pathway (*IL10* and *IL2/IL21*). No statistically significant interactions were observed in SNPs related to Crohn's disease.

In the LPS-sensing pathway, we observed statistically significant interaction for SNPs associated with rectal cancer risk in the *TLR4*, *TNFRSFB1*, and *CD14* genes, though the direction was not consistent for men or women (Supplementary Table 6). *IL6* and *IL18* both had SNPs with statistically significant interactions by sex associated with an increase in rectal cancer risk among women and a decrease in risk among men. There were no statistically significant interactions observed for SNPs in the mucin or Crohn's pathways.

We also stratified the gene- and pathway-level associations by sex to investigate any potential modifications and found there were notable differences among genes and pathways for both cancer sites. Stratified results for colon cancer risk are presented in Supplementary Table 7. The *TLR4* gene-level and Crohn's disease pathway-level results observed for all participants remained statistically significant only among men. Furthermore, men had a statistically significant interaction with gut permeability at the inflammation pathway level ($P_{unadjusted} = 0.043$). Women were observed to have multiple statistically significant biomarker interactions with interleukin-related genes, but no interaction was observed at the overall inflammation pathway level. For rectal cancer, women had a statistically significant interaction for CRP with the mucin pathway as well as statistically significant *IL6* interactions with both gut permeability and CRP biomarkers (Supplementary Table 8). The previously observed interaction of *NF-κB* and CRP in association with rectal cancer remained statistically significant for men but not women.

Last, we determined if there was heterogeneity among SNP-specific risk estimates between colon and rectal cancer sites. Several SNPs associated with colon cancer statistically significantly differed from the corresponding association with rectal cancer. Three SNPs (rs1554973, rs1927914, rs2149356; all $P_{heterogeneity} \leq 0.04$) were in *TLR4* and two (rs3828309, rs4921466; $P_{heterogeneity} < 0.02$) were SNPs related to Crohn's disease (Table 3). The only associated risk for rectal cancer that was statistically significantly different than the risk for colon cancer was for rs3024509 in *IL10* ($P_{heterogeneity} < 0.01$).

Discussion

Despite the important role the gut barrier is hypothesized to play in the development of CRC [7,35–39], few studies have thoroughly investigated genetic variants related to gut barrier function and subsequent immune response in relation to colon and/or rectal cancer risks. In this large prospective nested case-control study within the EPIC cohort, we investigated possible associations of these processes at the SNP, gene and pathway levels with colon and rectal cancer risk. Furthermore, we assessed whether these possible associations are modified by circulating biomarker levels as well as by sex and tumor site.

We identified SNPs in genes related to endotoxin/LPS sensing, mucin synthesis, inflammation, and Crohn's disease associated with circulating LPS, flagellin, and/or CRP levels. Within the genetic variation in the mucin pathway, *ABCB1* was associated with overall gut permeability and LPS while *MUC6/MUC2* was associated with all three gut barrier function biomarkers. *ABCB1* encodes ABCB1 transporters, which are located in the cell membrane of the epithelium in the intestine, aiding in barrier function by influencing the absorption, distribution, and excretion of compounds and exogenous toxic agents [40,41]. Both the *MUC6* and *MUC2* genes, located on chromosome 11 and highly homologous [42], encode the secretory mucins 6 and 2 proteins, which form the insoluble mucous barrier to protect the epithelium [23]. Lost or altered function of genes in the mucin pathway could lead to a less effective mucin barrier and a greater prevalence of pathogenic bacteria interacting with the epithelium and possibly translocating to the lamina propria and systemic bloodstream.

The activation of TLR4 signaling pathways by binding with LPS via LBP and CD14 can stimulate antibody production through the modification of B cell responses [24] and induces

Table 3. Statistically significant associations (unadjusted $P < 0.05$) of SNPs with colon and rectal cancer risks using an additive model, the EPIC study, 1992–2003.

Gene/SNP	Cases	Controls	OR (95% CI)[a]	P[b]	$P_{heterogeneity}$[c]
Colon cancer					
Endotoxins/LPS					
TLR4					
rs10116253	859	1349	0.84 (0.73–0.96)	*0.01*	*0.13*
rs11536889	870	1365	1.19 (1.01–1.40)	*0.04*	*0.19*
rs1554973	853	1343	0.79 (0.68–0.91)	*<0.01*	*0.03*
rs1927911	854	1345	0.83 (0.72–0.95)	*0.01*	*0.09*
rs1927914	860	1356	0.85 (0.75–0.97)	*0.02*	*0.04*
rs2149356	820	1284	0.83 (0.72–0.96)	*0.01*	*0.03*
rs2770146	718	1118	1.22 (1.05–1.41)	*0.01*	*0.08*
rs7034482	832	1312	1.17 (1.02–1.34)	*0.02*	*1.00*
rs7873784	865	1364	0.75 (0.63–0.89)	*<0.01*	*0.09*
Inflammation					
IL12B					
rs4921466	758	1216	0.69 (0.53–0.89)	*<0.01*	*0.43*
Crohn's disease[d]					
rs3828309	869	1369	1.18 (1.04–1.33)	*0.01*	*0.02*
rs762421	868	1367	0.80 (0.70–0.90)	*<0.01*	*0.02*
Rectal Cancer					
Endotoxins/LPS					
TLR4					
rs10759934	474	1286	0.85 (0.73–0.98)	*0.03*	*0.29*
TNFRSF1B					
rs816050	502	1368	0.81 (0.67–0.98)	*0.03*	*0.17*
Mucins					
MUC6/MUC2					
rs11601642	502	1366	1.18 (1.02–1.38)	*0.03*	*0.38*
Inflammation					
IL10					
rs3024509	496	1330	0.54 (0.38–0.77)	*<0.01*	*<0.01*
Crohn's disease[d]					
rs3197999	500	1355	0.82 (0.70–0.97)	*0.02*	*0.45*

[a]Unconditional logistic regression adjusted for age at blood collection, sex, and study center.
[b]Not statistically significant after BH method for multiple testing correction.
[c]P for heterogeneity by tumor site (colon vs. rectal).
[d]SNPs previously found in GWAS to be associated with Crohn's disease [62].

pro-inflammatory cytokine production [13]. The cytokine production triggers the release of CRP which is considered a first line component of innate immune defense against bacterial pathogens and endotoxemia [43,44]. In line with this process, we found and association between SNPs in the *LBP* and *LPS* tolerance genes and serum CRP levels.

In our study, multiple SNPs and overall gene-wide variation in *TLR4* were associated with colon cancer risk. Polymorphisms in *TLR4* have been previously implicated to play a role in cancer development with a large focus on CRC. Slattery *et al.* investigated 8 SNPs in *TLR4* and found an association for rs11536898 (OR = 0.50 [95% CI: 0.29-0.87]; AA vs. CA/CC) with colon cancer [17]. Although we did not include it in our analysis, rs11536898 is in LD with rs10116253 which we similarly observed to be associated with a lower risk of colon cancer (OR = 0.84 [95% CI: 0.73–0.96]). Several other meta-analyses and publications have analyzed *TLR4* polymorphisms yet concluded conflicting results for both statistical significance and the direction of the associations. These inconsistencies might be due to small sample sizes, heterogenous study populations, and limited overlap in selected SNPs [45–48]. While we observed statistically significant unadjusted *P*-values for *TLR4* SNPs associated with both colon and rectal cancer risks, there appears to be a suggestively stronger association for colon cancer. TLR4 has a complex role in the development and progression of cancer as it induces the NF-κB-mediated production and secretion of pro-inflammatory cytokines, which can contribute to colon cancer development and progression [29,30].

At the gene level, *ALOX5* was associated with colon cancer risk when interacting with both the overall gut permeability score and LPS alone. *ALOX5* encodes 5-lipoxygenase (5-LO) and can regulate cell death through inflammation and lipid peroxidation [49]. 5-LO has been observed to play an antitumorigenic role in various cancer cell lines [50–53]. Human monocytic cells incubated with LPS was associated with increased 5-LO expression relative to cells incubated without LPS.

Table 4. *P*-values of pathway- and gene-level associations with colon cancer risk and interactions with circulating gut barrier function and inflammation biomarkers, the EPIC study, 1992–2003.[a]

Pathway/gene	No. of SNPs	No. of SNPs retained after LD[b] pruning	Gene or pathway only	Gut permeability score interaction	LPS interaction	Flagellin interaction	CRP interaction
			P[b]	*P*[b]	*P*[b]	*P*[b]	*P*[b]
LPS pathway	78	66	*0.07*	*0.96*	*0.84*	*0.86*	*0.49*
CD14	5	3	*0.32*	*0.61*	*0.54*	*0.56*	*0.15*
LBP	21	16	*0.36*	*0.86*	*0.91*	*0.53*	*0.39*
LPS tolerance	3	3	*0.39*	*0.93*	*0.89*	*0.86*	*0.16*
LY96	8	8	*0.78*	*0.87*	*0.44*	*0.93*	*0.52*
MYD88	1	1	*0.37*	*0.58*	*0.95*	*0.34*	*0.84*
TLR4	19	15	***0.01***	*0.32*	*0.27*	*0.27*	*0.46*
TNFRSF1B	21	20	*0.64*	*0.70*	*0.21*	*0.39*	*0.29*
Mucins	66	57	*0.47*	*0.54*	*0.91*	*0.21*	*0.41*
ABCB1	21	18	*0.77*	*0.96*	*0.99*	*0.81*	*0.36*
CAMP/ZNF589	1	1	*0.06*	*0.74*	*0.63*	*0.93*	*0.31*
CD38	12	11	*0.52*	*0.41*	*0.48*	*0.43*	*0.07*
MUC1	1	1	*0.32*	*0.75*	*0.94*	*0.56*	*0.51*
MUC12	4	4	*0.60*	*0.71*	*0.64*	*0.65*	*0.79*
MUC17	5	3	*0.74*	*0.76*	*0.39*	*0.82*	*0.24*
MUC6/MUC2	22	19	*0.83*	*0.08*	*0.28*	***0.02***	*0.39*
Inflammation	133	97	*0.64*	*0.16*	*0.10*	*0.39*	*0.72*
ALOX5	22	11	*0.89*	***0.01***	***<0.01***	*0.19*	*0.21*
IFNG	7	4	*0.32*	*0.76*	*0.82*	*0.80*	*0.06*
IL10	13	8	*0.60*	*0.47*	*0.66*	*0.43*	*0.59*
IL10R	9	7	*0.95*	***0.04***	*0.24*	***0.03***	*0.54*
IL12A	1	1	*0.33*	*0.13*	*0.45*	*0.08*	*0.62*
IL12B	18	17	*0.07*	*0.70*	*0.55*	*0.67*	*0.56*
IL18	4	4	*0.50*	*0.08*	*0.18*	*0.09*	*0.63*
IL1A/IL1B	11	8	*0.87*	*0.25*	*0.18*	*0.38*	*0.68*
IL2/IL21	6	5	*0.51*	*0.40*	*0.16*	*0.43*	*0.62*
IL6	13	10	*1.00*	*0.75*	*0.67*	*0.94*	*0.47*
IL8	1	1	*0.07*	*0.49*	*0.76*	*0.41*	*0.28*
NFKB1	22	15	*0.82*	*0.24*	*0.31*	*0.42*	*0.39*
RELA (p65)	2	2	*0.80*	*0.97*	*0.77*	*0.65*	*0.62*
TNF	4	4	*0.95*	*0.46*	*0.50*	*0.43*	*0.55*
Crohn's disease[c]	15	15	***0.01***	*0.78*	*0.45*	*0.82*	*0.22*

[a]Adjusted for age at blood collection, sex, and study center.
[b]Not statistically significant after BH method for multiple testing correction.
[c]SNPs previously found in GWAS to be associated with Crohn's disease [62].
Bold values indicate statistical significant of *P* raw < 0.05.

IL10 was associated with rectal cancer risk at both the SNP and gene level. This is consistent with a similar analysis of selenoprotein and selenium pathways and CRC risk by Fedirko *et al.* where the association of *IL10* SNP rs3024509 and rectal cancer had an unadjusted *P*-value of 0.014 [54]. Although the previous study did not find *IL10* statistically significant at the gene level, the authors did not stratify the gene-level analyses by cancer site and the greater number of colon cancers might attenuate a possible association. Both Tsidilis *et al.* [55] and Macarthur *et al.* [56] found a statistically significant decreased CRC risk per *IL10-1082* locus alleles, consistent with our rectal cancer results. IL10 is an anti-inflammatory cytokine and negative regular of immunological responses [57] which may attenuate a chronic inflammatory state in the gut and thereby reduce the risk of tumorigenesis. Loss of IL10 is also associated with the development of inflammatory bowel diseases, which is a known risk factor for CRC [35].

At the pathway level, we found SNPs related to Crohn's disease to be associated with colon but not rectal cancer risk. Crohn's disease does not commonly involve the rectum and roughly 42% of all cases are colonic Crohn's, supporting our findings [58]. Crohn's disease is hypothesized to contribute to cancer development due to the chronic inflammatory changes at the affected site [33].

Sex differences were also observed in association with cancer subsite development. A previous study of a healthy lifestyle index (HLI; composed by potentially modifiable lifestyle factors) and CRC provided evidence that obesity explained the observed association between HLI and colon cancer for men but not for women [59]. Based on previous work, the authors hypothesized that inflammation and

Table 5. *P*-values of pathway- and gene-level associations with rectal cancer risk and interactions with circulating gut barrier function and inflammation biomarkers, the EPIC study, 1992–2003.[a]

Pathway/gene	No. of SNPs	No. of SNPs retained after pruning	Gene or pathway only	Gut permeability score interaction	LPS interaction	Flagellin interaction	CRP interaction
			P[b]	*P*[b]	*P*[b]	*P*[b]	*P*[b]
LPS pathway	78	66	*0.59*	*0.90*	*0.73*	*0.75*	*0.71*
CD14	5	3	*0.40*	*0.22*	*0.13*	*0.34*	*0.17*
LBP	21	16	*0.43*	*0.90*	*0.73*	*0.90*	*0.24*
LPS tolerance	3	3	*0.21*	*0.84*	*0.65*	*0.68*	*0.80*
LY96	8	8	*0.36*	*0.81*	*0.88*	*0.76*	*0.64*
MYD88	1	1	*0.59*	*0.97*	*0.75*	*0.71*	*0.70*
TLR4	19	15	*0.30*	*0.93*	*0.34*	*1.00*	*0.35*
TNFRSF1B	21	20	*0.45*	*0.45*	*0.84*	*0.14*	*0.45*
Mucins	66	57	*0.85*	*0.69*	*0.52*	*0.38*	*0.67*
ABCB1	21	18	*0.67*	*0.53*	*0.72*	*0.62*	*0.94*
CAMP/ZNF589	1	1	*0.22*	*0.85*	*0.37*	*0.64*	*0.12*
CD38	12	11	*0.48*	*0.20*	*0.44*	*0.05*	*0.43*
MUC1	1	1	*0.70*	*0.87*	*0.99*	*0.79*	*0.57*
MUC12	4	4	*0.69*	*0.78*	*0.49*	*0.70*	*0.75*
MUC17	5	3	*0.73*	*0.15*	*0.08*	*0.50*	*0.57*
MUC6/MUC2	22	19	*0.30*	*0.35*	*0.18*	*0.69*	*0.83*
Inflammation	133	97	*0.18*	*0.68*	*0.81*	*0.26*	*0.20*
ALOX5	22	11	*0.83*	*0.19*	*0.09*	*0.12*	*0.95*
IFNG	7	4	*0.35*	*0.92*	*0.97*	*0.70*	*0.18*
IL10	13	8	***0.01***	*0.58*	*0.42*	*0.53*	*0.08*
IL10R	9	7	*0.78*	*0.57*	*0.89*	*0.34*	*0.55*
IL12A	1	1	*0.84*	*0.21*	*0.36*	*0.22*	*0.60*
IL12B	18	17	*0.83*	*0.32*	*0.36*	*0.50*	*0.70*
IL18	4	4	*0.26*	*0.22*	*0.57*	*0.07*	*0.10*
IL1A/IL1B	11	8	*0.50*	*0.46*	*0.30*	*0.51*	*0.56*
IL2/IL21	6	5	*0.50*	*0.73*	*0.27*	*0.92*	*0.60*
IL6	13	10	*0.73*	*0.17*	*0.32*	*0.05*	*0.07*
IL8	1	1	*0.76*	*0.09*	*0.35*	*0.05*	*0.82*
NFKB1	22	15	*0.88*	*0.76*	*0.61*	*0.39*	***0.02***
RELA (p65)	2	2	*0.83*	*0.89*	*0.72*	*0.89*	*0.56*
TNF	4	4	*0.74*	*0.55*	*0.71*	*0.33*	*0.67*
Crohn's disease[c]	15	15	*0.33*	*0.71*	*0.57*	*0.46*	*0.99*

[a]Adjusted for age at blood collection, sex, and study center.
[b]Not statistically significant after BH method for multiple testing correction.
[c]SNPs previously found in GWAS to be associated with Crohn's disease [62].
P values less than 0.05 are bolded.

oxidative stress may underlie the association in men, whereas hyperinsulinemia may be the potential biological mechanism in women [59,60]. Consistent with this theory, previous studies found that men had higher LPS-induced TNF-α and IL-1b but not IL-6 [61]. We observed a statistically significant interaction by sex for the association of IL-6 and an increase in rectal cancer risk among women. In colon cancer, the genes with SNPs that showed different cancer risk associations between men and women were primarily related to gut barrier signaling pathways. If genetic variants in genes related to gut barrier integrity result in under expression of the resulting proteins in men leading to the translocation of LPS, this would result in higher levels of pro-inflammatory cytokines than women, supporting the inflammation hypothesis mentioned above. Furthermore, this could possibly explain our observations of a decrease in colon cancer risk per minor allele of *TNFRSF1B* and *LBP* SNPs in men, but an increase in women.

The strengths of our study include its prospective design, relatively large sample size, and high follow-up rate in the base cohort. In addition, the selection of pathways, genes, and SNPs was hypothesis driven. We used the detailed data from EPIC to address potential confounding by body size and other factors and our analyses suggested no or little confounding. However, we cannot altogether discount the possibility of residual confounding nor that changes in lifestyle habits between enrollment into the cohort and the eventual cancer diagnosis may have taken place. Although our study was large, the power was limited when conducting interaction and stratified analyses, especially by sex and tumor location. Due to unavailable data, we were not able to account for more detailed tumor molecular profiling such as microsatellite instability.

We were also limited by insufficient power to consider more specific tumor locations within the colon and rectum. In addition, none of our results retained statistical significance after BH correction for multiple testings. As we were limited by the published literature on gut barrier and inflammation genes at the time of gene and pathway selection for genotyping, we may not have included all relevant genes.

In conclusion, this study has identified possible genetic variations in gut barrier function and inflammatory genes and pathways associated with CRC risk in European populations. Repeating this investigation in a larger population is needed to determine if these findings are replicable. In addition, experimental studies are needed to understand possible biological mechanisms of the suggestive associations.

Supplementary data

Supplementary Tables 1-8 are available at *Mutagenesis* Online.

Acknowledgements

The authors would like to thank B. Hemon for his assistance in database preparation. Many thanks to the National Institute for Public Health and the Environment (RIVM), Bilthoven and Utrecht, The Netherlands, EPIC Asturias, and EPIC-Ragusa, Hyblean Association for Epidemiological Research, AIRE—ONLUS for their contribution and ongoing support to the EPIC Study. We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research.

Conflict of Interest

The authors have no relevant conflicts of interest to disclose.

Funding

V. Fedirko is supported by the Cancer Prevention and Research Institute of Texas (CPRIT) Rising Stars Award (grant number RR200056). D.J. Hughes is supported by the Health Research Board (HRB) of Ireland health research award HRA-POR-2013-397. M. Jenab was funded for this work by Wereld Kanker Onderzoek Fonds as part of the World Cancer Research Fund (WCRF) International Regular Grant Programme (grant number 2010–251). D.J. Hughes and M. Jenab are supported by the EU Marie Curie Doctoral Training Network "Colomark" (https://www.colomark.org/). The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS)—Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology—ICO (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford) (UK). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

1. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;**486**:207–14.
2. Lazar V, Ditu L-M, Pircalabioru GG, *et al*. Aspects of gut microbiota and immune system interactions in infectious diseases, immunopathology, and cancer. *Front Immunol* 2018;**9**:1830:1–18.
3. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* 2012;**10**:323–35.
4. Ramos HC, Rumbo M, Sirard J-C. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends Microbiol* 2004;**12**:509–17.
5. Celli JP, Turner BS, Afdhal NH, *et al*. *Helicobacter pylori* moves through mucus by reducing mucin viscoelasticity. *Proc Natl Acad Sci USA* 2009;**106**:14321–6.
6. Bischoff SC, Barbara G, Buurman W, *et al*. Intestinal permeability—a new target for disease prevention and therapy. *BMC Gastroenterol* 2014;**14**:189.
7. Bhatt AP, Redinbo MR, Bultman SJ. The role of the microbiome in cancer development and therapy. *CA Cancer J Clin* 2017;**67**:326–44.
8. Mohammad S, Thiemermann C. Role of metabolic endotoxemia in systemic inflammation and potential interventions. *Front Immunol* 2021;**11**:594150:1–16.
9. Newton K, Dixit VM. Signaling in innate immunity and inflammation. *Cold Spring Harb. Perspect. Biol* 2012;**4**:a006049.
10. Singh N, Baby D, Rajguru JP, *et al*. Inflammation and cancer. *Ann Afr Med* 2019;**18**:121–6.
11. Ahmad Kendong SM, Raja Ali RA, Nawawi KNM, *et al*. Gut dysbiosis and intestinal barrier dysfunction: potential explanation

for early-onset colorectal cancer. *Front Cell Infect Microbiol* 2021;**11**:744606.
12. Farhana A, Khan YS. Biochemistry, Lipopolysaccharide. In *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2023. https://www-ncbi-nlm-nih-gov.proxy.library.emory.edu/books/NBK554414/
13. Lu Y-C, Yeh W-C, Ohashi PS. LPS/TLR4 signal transduction pathway. *Cytokine* 2008;**42**:145–51.
14. Solt LA, May MJ. The IkappaB kinase complex: master regulator of NF-kappaB signaling. *Immunol Res* 2008;**42**:3–18.
15. Mitchell S, Vargas J, Hoffmann A. Signaling via the NFκB system. *Wiley Interdiscip Rev Syst Biol Med* 2016;**8**:227–41.
16. Wang S, Liu Z, Wang L, *et al*. NF-kappaB signaling pathway, inflammation and colorectal cancer. *Cell Mol Immunol* 2009;**6**:327–34.
17. Slattery ML, Herrick JS, Bondurant KL, *et al*. Toll-like receptor genes and their association with colon and rectal cancer development and prognosis. *Int J Cancer* 2012;**130**:2974–80.
18. Omrane I, Baroudi O, Kourda N, *et al*. Positive link between variant Toll-like receptor 4 (Asp299Gly and Thr399Ile) and colorectal cancer patients with advanced stage and lymph node metastasis. *Tumour Biol*2014;**35**:545–51.
19. Pimentel-Nunes P, Teixeira AL, Pereira C, *et al*. Functional polymorphisms of Toll-like receptors 2 and 4 alter the risk for colorectal carcinoma in Europeans. *Dig Liver Dis* 2013;**45**:63–9.
20. Li X-X, Sun G-P, Meng J, *et al*. Role of toll-like receptor 4 in colorectal carcinogenesis: a meta-analysis. *PLoS One* 2014;**9**:e93904.
21. Doubeni CA, Major JM, Laiyemo AO, *et al*. Contribution of behavioral risk factors and obesity to socioeconomic differences in colorectal cancer incidence. *J Natl Cancer Inst* 2012;**104**:1353–62.
22. Wolin KY, Yan Y, Colditz GA, *et al*. Physical activity and colon cancer prevention: a meta-analysis. *Br J Cancer* 2009;**100**:611–6.
23. Jensen OM. Different age and sex relationship for cancer of subsites of the large bowel. *Br J Cancer* 1984;**50**:825–9.
24. Riboli E, Hunt KJ, Slimani N, *et al*. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;**5**:1113–24.
25. Bingham S, Riboli E. Diet and cancer—the European Prospective Investigation into Cancer and Nutrition. *Nat Rev Cancer* 2004;**4**:206–15.
26. Jenab M, Bueno-de-Mesquita HB, Ferrari P, *et al*. Association between pre-diagnostic circulating vitamin D concentration and risk of colorectal cancer in European populations: a nested case-control study. *BMJ* 2010;**340**:b5500.
27. Ziegler TR, Luo M, Estívariz CF, *et al*. Detectable serum flagellin and lipopolysaccharide and upregulated anti-flagellin and lipopolysaccharide immunoglobulins in human short bowel syndrome. *Am J Physiol Regul Integr Comp Physiol* 2008;**294**:R402–10.
28. Sitaraman SV, Klapproth J-M, Moore DA, *et al*. Elevated flagellin-specific immunoglobulins in Crohn's disease. *Am J Physiol Gastrointest Liver Physiol* 2005;**288**:G403–406.
29. Kong SY, Tran HQ, Gewirtz AT, *et al*. Serum endotoxins and flagellin and risk of colorectal cancer in the European prospective investigation into cancer and nutrition (EPIC) cohort. *Cancer Epidemiol Biomarkers Prev* 2016;**25**:291–301.
30. Aleksandrova K, Jenab M, Boeing H, *et al*. Circulating C-reactive protein concentrations and risks of colon and rectal cancer: a nested case-control study within the European Prospective Investigation into Cancer and Nutrition. *Am J Epidemiol* 2010;**172**:407–18.
31. Ramagopalan SV, Heger A, Berlanga AJ, *et al*. A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. *Genome Res* 2010;**20**:1352–60.
32. Barrett JC, Hansoul S, Nicolae DL, *et al*; NIDDK IBD Genetics Consortium. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008;**40**:955–62.
33. Fedirko V, Mandle HB, Zhu W, *et al*. Vitamin D-related genes, blood vitamin D levels and colorectal cancer risk in western European populations. *Nutrients* 2019;**11**:1954.
34. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 1995;**57**:289–300.
35. Quaglio AEV, Grillo TG, De Oliveira ECS, *et al*. Gut microbiota, inflammatory bowel disease and colorectal cancer. *World J Gastroenterol* 2022;**28**:4053–60.
36. Bai X, Wei H, Liu W, *et al*. Cigarette smoke promotes colorectal cancer through modulation of gut microbiota and related metabolites. *Gut* 2022;**71**:2439–50.
37. Yang J, Wei H, Zhou Y, *et al*. High-fat diet promotes colorectal tumorigenesis through modulating gut microbiota and metabolites. *Gastroenterology* 2022;**162**:135–149.e2.
38. Brennan CA, Garrett WS. Gut microbiota, inflammation, and colorectal cancer. *Annu Rev Microbiol* 2016;**70**:395–411.
39. Genua F, Raghunathan V, Jenab M, *et al*. The role of gut barrier dysfunction and microbiome dysbiosis in colorectal cancer development. *Front Oncol* 2021;**11**:626349.
40. Andersen V, Svenningsen K, Knudsen LA, *et al*. Novel understanding of ABC transporters ABCB1/MDR/P-glycoprotein, ABCC2/MRP2, and ABCG2/BCRP in colorectal pathophysiology. *World J Gastroenterol* 2015;**21**:11862–76.
41. Hodges LM, Markova SM, Chinn LW, *et al*. Very important pharmacogene summary: ABCB1 (MDR1, P-glycoprotein). *Pharmacogenet Genomics* 2011;**21**:152–61.
42. McGuckin MA, Thornton DJ, Whitsett JA. Chapter 14–mucins and mucus. In: Mestecky J *et al.* (eds.). *Mucosal Immunology (Fourth Edition)*, Boston: Academic Press, 2015, 231–250. https://doi.org/10.1016/B978-0-12-415847-4.00014-8
43. Volanakis JE. Human C-reactive protein: expression, structure, and function. *Mol Immunol* 2001;**38**:189–97.
44. Jenny NS, Cushman M. C-Reactive protein. *Circ Res* 2014;**114**:596–7.
45. Moaaz M, Youssry S, Moaz A, *et al*. Study of toll-like receptor 4 gene polymorphisms in colorectal cancer: correlation with clinicopathological features. *Immunol Invest* 2020;**49**:571–84.
46. Zhu L, Wang Y, Jie G, *et al*. Association between Toll-like receptor 4 and interleukin 17 gene polymorphisms and colorectal cancer susceptibility in Northeast China. *Med Oncol* 2014;**31**:73.
47. Davoodi H, Seow HF. Variant Toll-like receptor4 (Asp299Gly and Thr399Ile alleles) and Toll-like receptor2 (Arg753Gln and Arg677Trp alleles) in colorectal cancer. *Iran J Allergy Asthma Immunol* 2011;**10**:91–9.
48. Sheng WY, Yong Z, Yun Z, *et al*. Toll-like receptor 4 gene polymorphisms and susceptibility to colorectal cancer: a meta-analysis and review. *Arch Med Sci*2015;**11**:699–707.
49. Sun Q-Y, Zhou H-H, Mao X-Y. Emerging roles of 5-lipoxygenase phosphorylation in inflammation and cell death. *Oxid Med Cell Longev* 2019;**2019**:e2749173.
50. Wang Y, Skibbe JR, Chao H, *et al*. ALOX5 exhibits anti-tumor and drug-sensitizing effects in MLL-rearranged leukemia. *Sci Rep* 2017;**7**(1):1853.
51. Poczobutt JM, Nguyen TT, Hanson D, *et al*. Deletion of 5-lipoxygenase in the tumor microenvironment promotes lung cancer progression and metastasis through regulating T cell recruitment. *J Immunol (Baltimore, Md.: 1950)* 2016;**196**:891–901.
52. Sarveswaran S, Chakraborty D, Chitale D, *et al*. Inhibition of 5-lipoxygenase selectively triggers disruption of c-Myc signaling in prostate cancer cells. *J Biol Chem* 2015;**290**:4994–5006.
53. Zhou G-X, Ding X-L, Wu S-B, *et al*. Inhibition of 5-lipoxygenase triggers apoptosis in pancreatic cancer cells. *Oncol Rep* 2015;**33**:661–8.
54. Fedirko V, Mandle HB, Zhu W, *et al*. Association of selenoprotein and selenium pathway genotypes with risk of colorectal cancer and interaction with selenium status. *Nutrients* 2019;**11**:1954.
55. Tsilidis KK, Helzlsouer KJ, Smith MW, *et al*. Association of common polymorphisms in IL10, and in other genes related to inflammatory response and obesity with colorectal cancer. *Cancer Causes Control* 2009;**20**:1739–51.

56. Macarthur M, Sharp L, Hold GL, *et al*. The role of cytokine gene polymorphisms in colorectal cancer and their interaction with aspirin use in the Northeast of Scotland. *Cancer Epidemiol Biomarkers Prev* 2005;**14**:1613–8.
57. Rutz S, Ouyang W. Regulation of interleukin-10 expression. *Adv Exp Med Biol* 2016;**941**:89–116.
58. Arms-Williams B, Hawthorne AB, Cannings-John R, *et al*. Changes in incidence and clinical features of inflammatory bowel disease in Cardiff, UK over 50 years: an update for 2005–2016. *Scand J Gastroenterol* 2023;**58**:619–26.
59. Aleksandrova K, Pischon T, Jenab M, *et al*. Combined impact of healthy lifestyle factors on colorectal cancer: a large European cohort study. *BMC Med* 2014;**12**:1–15.
60. Aleksandrova K, Drogan D, Boeing H, *et al*. Adiposity, mediating biomarkers and risk of colon cancer in the European prospective investigation into cancer and nutrition study. *Int J Cancer* 2014;**134**:612–21.
61. Imahara SD, Jelacic S, Junker CE, *et al*. The influence of gender on human innate immunity. *Surgery* 2005;**138**:275–82.
62. GWAS Catalog. https://www.ebi.ac.uk/gwas/