University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Bouras, E;Yu, R;Kim, AE;Markozannes, G;Murphy, N;Albanes, D;Anderson, LN;Barry, EL;Berndt, SI;Bishop, DT;Brenner, H;Burnett-Hartman, A;Campbell, PT;Carreras-Torres, R;Chan, AT;Cheng, I;Devall, MA;Diez-Obrero, V;Dimou, N;Drew, DA;Gruber, SB;Gsur, A;Hoffmeister, M;Hsu, L;Huyghe, JR;Kawaguchi, E;Keku, TO;Kundaje, A;Küry, S;Le Marchand, L;Lewinger, JP;Li, L;Lynch, BM;Moreno, V;Morrison, JL;Newton, CC;Obón-Santacana, M;Palmer, JR;Papadimitriou, N;Pellatt, AJ;Peoples, AR;Pharoah, PDP;Platz, EA;Qu, C;Ruiz-Narvaez, E;Mendez, JS;Schoen, RE;Stern, MC;Thomas, CE;Tian, Y;Um, CY;Visvanathan, K;Vodicka, P;Vymetalkova, V;White, E;Wolk, A;Woods, MO;Wu, AH;Gunter, MJ;Gauderman, WJ;Peters, U;Evangelou, M;Tsilidis, KK

Title:
Using gene-environment interactions to explore pathways for colorectal cancer risk

Date:
2025-11-01

Citation:
Bouras, E., Yu, R., Kim, A. E., Markozannes, G., Murphy, N., Albanes, D., Anderson, L. N., Barry, E. L., Berndt, S. I., Bishop, D. T., Brenner, H., Burnett-Hartman, A., Campbell, P. T., Carreras-Torres, R., Chan, A. T., Cheng, I., Devall, M. A., Diez-Obrero, V., Dimou, N. ,... Tsilidis, K. K. (2025). Using gene-environment interactions to explore pathways for colorectal cancer risk. Ebiomedicine, 121, pp.105964-. https://doi.org/10.1016/j.ebiom.2025.105964.

Persistent Link:
https://hdl.handle.net/11343/367726

License:
CC BY

Using gene-environment interactions to explore pathways for colorectal cancer risk

Emmanouil Bouras,[a,b,*] Ren Yu,[c] Andre E. Kim,[d] Georgios Markozannes,[a,b] Neil Murphy,[e] Demetrius Albanes,[f] Laura N. Anderson,[g] Elizabeth L. Barry,[h] Sonja I. Berndt,[f] D. Timothy Bishop,[i] Hermann Brenner,[j,k,l] Andrea Burnett-Hartman,[m] Peter T. Campbell,[n] Robert Carreras-Torres,[o,p] Andrew T. Chan,[q,r,s,t] Iona Cheng,[u] Matthew A. Devall,[v,w] Virginia Diez-Obrero,[o,x,y,z] Niki Dimou,[e] David A. Drew,[s,t] Stephen B. Gruber,[aa] Andrea Gsur,[ab] Michael Hoffmeister,[j] Li Hsu,[ac,ad] Jeroen R. Huyghe,[ac] Eric Kawaguchi,[d] Temitope O. Keku,[ae] Anshul Kundaje,[af,ag] Sébastien Küry,[ah] Loïc Le Marchand,[ai] Juan Pablo Lewinger,[d] Li Li,[aj] Brigid M. Lynch,[ak,al] Victor Moreno,[o,x,y,z] John L. Morrison,[d] Christina C. Newton,[am] Mireia Obón-Santacana,[o,x,y] Julie R. Palmer,[an] Nikos Papadimitriou,[e] Andrew J. Pellatt,[ao] Anita R. Peoples,[am] Paul D. P. Pharoah,[ap] Elizabeth A. Platz,[aq] Conghui Qu,[ac] Edward Ruiz-Narvaez,[ar] Joel Sanchez Mendez,[as] Robert E. Schoen,[at] Mariana C. Stern,[as] Claire E. Thomas,[ac] Yu Tian,[au,av] Caroline Y. Um,[am] Kala Visvanathan,[aq] Pavel Vodicka,[aw,ax,ay] Veronika Vymetalkova,[aw,ax,ay] Emily White,[ac,az] Alicja Wolk,[ba] Michael O. Woods,[bb] Anna H. Wu,[as] Marc J. Gunter,[a,e] W. James Gauderman,[d,bc] Ulrike Peters,[ac,az,bc] Marina Evangelou,[c,bc] and Konstantinos K. Tsilidis[a,b,**,bc]

[a]Department of Epidemiology and Biostatistics, School of Public Health—Faculty of Medicine, Imperial College London, London, UK
[b]Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece
[c]Department of Mathematics, Faculty of Natural Sciences, Imperial College London, London, UK
[d]Division of Biostatistics, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[e]Nutrition and Metabolism Branch, International Agency for Research on Cancer, Lyon, France
[f]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
[g]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
[h]Department of Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, NH, USA
[i]Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK
[j]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
[k]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
[l]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
[m]Institute for Health Research, Kaiser Permanente Colorado, Denver, CO, USA
[n]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA
[o]Colorectal Cancer Group, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, 08908, Barcelona, Spain
[p]Digestive Diseases and Microbiota Group, Girona Biomedical Research Institute (IDIBGI), Salt, 17190, Girona, Spain
[q]Broad Institute of Harvard and MIT, Cambridge, MA, USA
[r]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[s]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[t]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[u]Department of Epidemiology and Biostatistics, University of California-San Francisco, San Francisco, CA, USA
[v]Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA
[w]Department of Public Health Sciences, Center for Public Health Genomics, Charlottesville, VA, USA
[x]Unit of Biomarkers and Susceptibility (UBS), Oncology Data Analytics Program (ODAP), Catalan Institute of Oncology (ICO), Barcelona, 08908, Spain
[y]Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), 28029, Madrid, Spain
[z]Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, 08908, Spain
[aa]Department of Medical Oncology & Therapeutics Research and Center for Precision Medicine, City of Hope National Medical Center, Duarte, CA, USA
[ab]Center for Cancer Research, Medical University of Vienna, Vienna, Austria
[ac]Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA
[ad]Department of Biostatistics, University of Washington, Seattle, WA, USA
[ae]Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, NC, USA
[af]Department of Computer Science, Stanford University, Stanford, CA, USA

*Corresponding author. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, White City Campus, 90 Wood Lane, London W12 0BZ, UK.
**Corresponding author. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, White City Campus, 90 Wood Lane, London W12 0BZ, UK.
E-mail addresses: ebouras@imperial.ac.uk (E. Bouras), k.tsilidis@imperial.ac.uk (K.K. Tsilidis).
[bc]These authors contributed equally and are joint last.

[ag]Department of Genetics, Stanford University, Stanford, CA, USA
[ah]Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) Nantes, Nantes, France
[ai]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA
[aj]Department of Family Medicine, University of Virginia, Charlottesville, VA, USA
[ak]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia
[al]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
[am]Department of Population Science, American Cancer Society, Atlanta, GA, USA
[an]Slone Epidemiology Center, at Boston University, Boston, MA, USA
[ao]Intermountain Health, Salt Lake City, UT, USA
[ap]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
[aq]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[ar]Department of Nutritional Sciences, University of Michigan School of Public Health, Ann Arbor, MI, USA
[as]Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[at]Departments of Medicine and Epidemiology, University of Pittsburgh, Pittsburgh, PA, USA
[au]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
[av]School of Public Health, Capital Medical University, Beijing, China
[aw]Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic
[ax]Faculty of Medicine and Biomedical Center in Pilsen, Charles University, Pilsen, Czech Republic
[ay]Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic
[az]Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA
[ba]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
[bb]Memorial University of Newfoundland, Discipline of Genetics, St. John's, Canada

eBioMedicine
2025;121: 105964
Published Online xxx
https://doi.org/10.1016/j.ebiom.2025.105964

Summary

Background Colorectal cancer (CRC) is a significant public health concern, highlighting the critical need for identifying novel intervention targets for its prevention.

Methods We conducted genome-wide interaction analyses for 15 exposures with established or putative CRC risk [body mass index (BMI), height, physical activity, smoking, type 2 diabetes, use of menopausal hormone therapy, non-steroidal anti-inflammatory drugs, and intake of alcohol, calcium, fibre, folate, fruits, processed meat, red meat, and vegetables], and used interaction estimates to explore pathways and genes underlying CRC risk. The adaptive combination of Bayes Factors (ADABF), and over-representation analysis (ORA) were used for pathway analyses, and findings were further investigated using publicly available resources [hallmarks of cancer, Open Targets Platform (OTP)].

Findings A total of 1973 pathways using ADABF, and 840 pathways using ORA, out of the 2950 analysed, were enriched ($P < 0.05$) for at least one exposure, as well as 1227 genes within the enriched pathways. Data were available for 811/1227 coding genes in the OTP, 241 of which were supported by strong relative abundance of prior evidence (overall OTP score > 0.05). Fifty percent of the genes (617/1227) mapped to at least one hallmark of cancer, most of which (388/617) pertained to the Sustaining Proliferative Signalling hallmark. Our findings reflect previously established pathways for CRC risk and highlight the emerging importance of several less studied genes. Common pathways were found for several combinations of exposures, potentially suggesting common underlying mechanisms.

Interpretation The results of the present analysis provide a basis for further functional research. If confirmed, they may help elucidate the etiological associations between risk factors and CRC risk and ultimately inform personalized prevention strategies.

Funding This study was funded by *Cancer Research UK* (CRUK; *grant number: PPRCPJT\100005*) and World Cancer Research Fund International (WCRF; *IIG_FULL_2020_022*). Funding for grant IIG_FULL_2020_022 was obtained from Wereld Kanker Onderzoek Fonds (WKOF) as part of the World Cancer Research Fund International grant programme. Full funding details for the individual consortia are provided in the acknowledgements.

Copyright © 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords: Colorectal cancer; Gene-environment interactions; Molecular pathways; Mechanisms; Pathway analysis

Research in Context

Evidence before this study

Gene-environment interaction analyses have identified several associations and mechanisms related to colorectal cancer (CRC) risk. However, they typically focus on a limited set of genetic loci that remain statistically significant after correction for multiple comparisons, leaving potentially suggestive associations underexploited.

Added value of this study

We derived gene-environment interaction estimates for 15 CRC risk factors (e.g., BMI, smoking, diet, medications) across genes mapped to 2950 biological pathways, and conducted pathway analyses, to highlight genes and pathways for CRC risk. The relevance of the identified genes and pathways to CRC was further assessed using publicly available resources. Our findings support the involvement of several genes and pathways, many of which have been relatively understudied in the context of CRC. Notably, we identified pathways and genes that were enriched across multiple exposures, suggesting the presence of shared underlying mechanisms.

Implications of all the available evidence

The results of the present analysis provide a basis for further functional research. If confirmed, they may help elucidate the etiological associations between risk factors and CRC risk and ultimately inform personalized prevention strategies.

Introduction

Colorectal cancer (CRC) is the third most common type of cancer globally, and the second cause of cancer death, with an estimate of approximately two million incident cases and one million deaths in 2020.[1] There has been an upward trend over the past decades, with projections of incidence cases estimated to increase by more than 60% by 2040.[2] Such an increase is largely driven by growth of the older population, but also reflects changes in the prevalence of major CRC risk factors, such as smoking, adiposity, and physical inactivity,[3] especially among younger adults in prescreening ages (<50).[4]

Investigating the interactions between genome-wide variants and relevant exposures for risk of CRC is useful in identifying novel risk loci, but also offers insights into the biological pathways that are implicated in the development of CRC.[5] Such gene-environment interaction analyses have yielded several associations for CRC risk (and potentially mechanisms), such as in loci 6p21.33 [Activating Transcription Factor 6 Beta (*ATF6B*)] and 3p12.1 [Cell Adhesion Molecule 2 (*CADM2*)] as interaction loci with smoking, 10q24.2 [Cytochrome C Oxidase Assembly Homolog COX15 (*COX15*)] with alcohol consumption, 12p13.1 [Glutamate Ionotropic Receptor NMDA Type Subunit 2B (*GRIN2B*)] and 6q22.1 [Discoidin, CUB And LCCL Domain Containing 1 (*DCBLD1*)] with menopausal hormone therapy (MHT) use, 8q24.11 [Solute Carrier Family 30 Member 8 (*SLC30A8*)] with diabetes, 7q31.1 [Solute Carrier Family 26 Member 3 (*SLC26A3*)] with fibre, 1p31.1 [Neuronal Growth Regulator 1 (*NEGR1*)] with fruit consumption, 8q24.13 [Hyaluronan Synthase 2 (*HAS2*)] and 18q21.1 [SMAD Family Member 7 (*SMAD7*)] with red meat consumption, 3p25.2 [Synapsin II (*SYN2*)/tissue inhibitor of metalloproteinase 4 (*TIMP4*)] with folate, 15q13.3 [Formin 1/Gremlin 1 (FMN1/GREM1)] with body mass index (BMI).[6–13] These high-throughput analyses entail a wide genetic coverage but typically focus on the limited set of genetic loci that remain statistically significant after correction for multiple comparisons, leaving a wealth of information (including potentially suggestive associations) inevitably underexploited. A holistic investigation of those associations may shed light on novel pathways for CRC risk and identify targets for CRC prevention, which is of critical importance due to the rising incidence.

In this study, we aimed to investigate biological pathways linked to CRC development by conducting genetic pathway-environment interaction analyses. Using estimates from genome-wide interaction studies (GWIS) of 15 exposures with established or putative CRC risk, we examined associated and over-represented genes and pathways in a pooled dataset of 36 studies including approximately 85,000 participants of European ancestry.

Methods

The overview of the project is presented in Fig. 1.

Study participants

Details on the study design and the GWIS are described elsewhere.[6–13] In brief, we included data from a total of 36,415 CRC cases and 48,451 controls from 36 studies contributing to three consortia—the multi-centred Colon Cancer Family Registry Cohort (CCFRC), the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), and the Colorectal Transdisciplinary Study (CORECT) (S.Table 1). Cases were confirmed by medical records, pathological reports, or death-certificate information, and controls were matched on age, sex, and enrolment date/trial group,

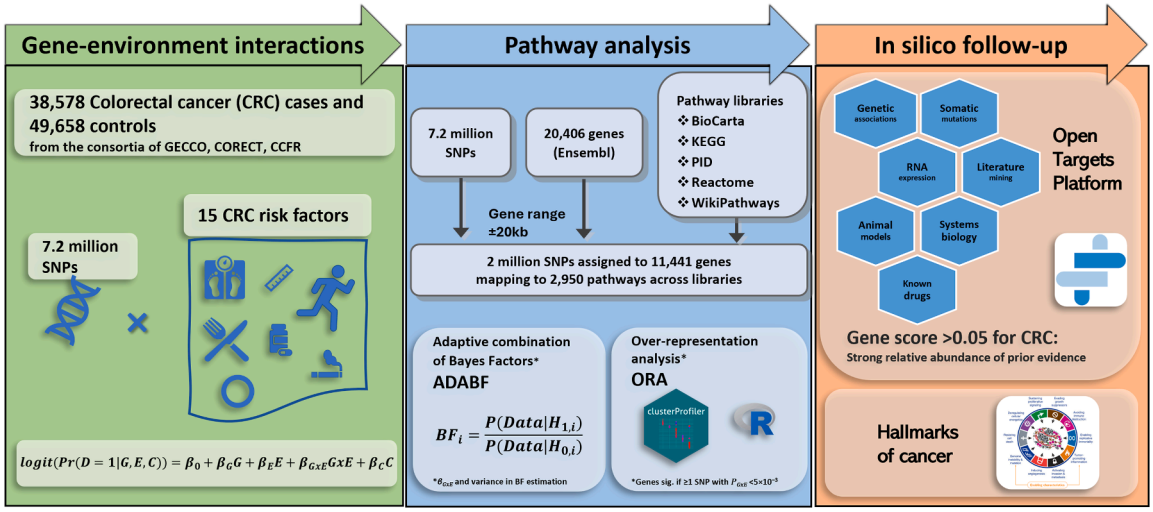


Fig. 1: Project overview. ADABF: adaptive combination of Bayes factors; BF: Bayes factors; CCFR: Colon Cancer Family Registry Cohort; CRC: Colorectal cancer; GECCO: Genetics and Epidemiology of Colorectal Cancer Consortium; KEGG: Kyoto Encyclopedia of Genes and Genomes; ORA: Over-representation analysis; PID: Pathway Interaction Database; SNPs: Single nucleotide polymorphisms.

when applicable. All participants gave written informed consent and studies were approved by their respective Institutional Review Boards/ethnic committees. This investigation was restricted to European-ancestry participants.

Exposures

Information on exposures and other variables of interest, such as demographics and environmental risk factors, was obtained using structured questionnaires and/or interviews, and a multi-step data-harmonization procedure was carried out at the GECCO coordinating centre (Fred Hutchinson Cancer Center), to reconcile each study's protocols and data-collection instruments.[14] We used 15 exposures (27 separate variables), namely alcohol, calcium, fibre, folate, fruit, processed meat, red meat, and vegetable intake, BMI, height, physical activity, smoking, type 2 diabetes (T2D), MHT, non-steroidal anti-inflammatory drugs (NSAIDs), and several sub-categories [e.g., smoking was assessed as a categorical variable (ever vs. never smoker) and as a continuous variable (pack-years of smoking)] (S.Table 2). These exposures were prioritized for analysis based on prior evidence of their relevance to colorectal cancer and have been systematically harmonized across participating studies within the GECCO consortium. Exposure specific descriptives and sample sizes are presented in S.Table 3.

Genotyping, quality assurance/control and imputation

Details on quality control and genotyping have been previously described.[15,16] Briefly, genotyping call rates (<97%), heterozygosity, duplicates or next of kin individuals, and discrepancies between self-reported and genotypic sex, were used as criteria to exclude study participants, that were further limited to only European ancestry, as determined by self-reported race and principal components clustering with 1000 Genomes EUR super-populations.[17] Genetic markers were excluded based on missing call rates (>2–5%), departures from Hardy–Weinberg Equilibrium (HWE) (P-value $<1 \times 10^{-4}$), and discordant genotype calls within duplicate samples. Genotypes were imputed to the Haplotype Reference Consortium (HRC version r1.1) using the University of Michigan Imputation Server and filtered out imputed single nucleotide polymorphisms (SNPs) based on imputation accuracy ($R^2 > 0.8$) and minor allele frequency (MAF) > 1%, leaving a total of more than 7.2 million SNPs that were used in subsequent interaction analyses. Principal component analysis (PCA) for population stratification assessment was performed using PLINK1.9.[15,16]

Statistics

For each exposure, gene-environment (GxE) interaction scans were performed using the R software package *GxEScanR* developed by members of our team.[18] These included traditional 1 degree of freedom analyses to test the multiplicative scale interaction (GxE) using logistic regression models, adjusting for age, sex, and the first three principal components of ancestry, and total energy (only for dietary exposures).

Only SNPs that map onto protein-coding genes within the pathway libraries of interest were brought forward to the pathway analyses. Each SNP was mapped to a protein-encoding gene extended by 20 kb [gene start –20 kb; gene end + 20 kb], using the genomic

coordinates obtained from Ensembl's BioMart (assessed: Jan 16, 2022).[19,20] This process resulted in approximately 2.8 million SNPs mapping to proteins coding genes. Of these, approximately 2 million SNPs were assigned to 11,441 protein-encoding genes (S. Figure 1), of 2950 canonical pathways described in the gene-set (pathway) libraries of BIOCARTA (n = 292 gene-sets), Kyoto Encyclopedia of Genes and Genomes (KEGG) (n = 186), Pathway Interaction Database (PID) (n = 196), REACTOME (n = 1614) and WikiPathways (n = 662) from the Molecular Signatures Database (MSigDB).[21–23] Details on the different gene-set libraries used can be found in the MSigDB resource and are briefly described in the Methods Supplement.[21–23]

We employed two complementary approaches in parallel, each with distinct methodological characteristics, for identifying associated and over-represented pathways linked to the development of CRC: the adaptive combination of Bayes Factors (ADABF) method, and over-representation analysis (ORA). The ADABF tests the association null hypothesis: "*no pathway gene-environment interactions are associated with CRC*", which is weaker than the competitive null hypothesis typically tested in the ORA: "*the pathway gene-environment interactions are more associated with CRC than genes outside the pathway*".[24] The two approaches are described below.

We used the ADABF method proposed by Lin et al., to aggregate the variants' information within genes and pathways.[24] In brief, the method uses a four-step procedure to compute adjusted *P*-values, for each gene/pathway. In the first step, approximate Bayes factors (BF) for all the SNPs in the gene/pathway of interest for L variants (SNPs) on a region (gene locus), are estimated: $BF = \sqrt{\frac{\widehat{V}}{\widehat{V}+W}}\exp\left(\frac{\widehat{\beta}W}{2\widehat{V}(\widehat{V}+W)}\right)$, where $\widehat{\beta}$ and $\widehat{V}$, are the beta and its corresponding variance from the interaction models, and W is the prior variance (here we used $W = 0.2^2 = 0.04$ as proposed previously[25]). The computed Bayes factors are then ordered as $BF_{(1)} \leq BF_{(2)} \leq \ldots \leq BF_{(L)}$, where the k-th smallest Bayes factor is denoted as $BF_{(k)}$. Based on the ordered Bayes factors, the summary scores (S) are estimated as: $S_k = \sum_{i=1}^{L} I(BF_{(i)} \geq BF_{(k)})\log(BF_{(i)}), k = 1, \ldots, L$, where $I(\cdot)$ is the indicator function. In our work we have followed the sequential resampling procedure described by Liu et al. for assessing the significance of association between each pathway and disease, with the number of sets drawn set to B = 1000. At the end of the procedure, *P*-values of association are computed and reported in the manuscript.[24]

We also performed an ORA to investigate whether genes from pre-defined pathways were presented more than expected in the list of genes per exposure. For each pathway, ORA generates the *P*-value of a hypergeometric test, where the alternative hypothesis indicates that the pathway contains more genes from the gene list (which is formed with the genes that have at least one SNP with *P*-value $<5 \times 10^{-3}$) than expected, using the following equation: $p = \sum_{j=n}^{K} \frac{\binom{L-M}{K-j}\binom{M}{j}}{\binom{L}{K}}$, where K is the number of genes in the gene list; L is the number of genes in the reference set of genes (background), M is the number of genes that are annotated to a particular gene-set (pathway), and n is the number of genes in the gene list mapping to M. The number of SNPs and genes, per exposure, with *P*-value $<5 \times 10^{-3}$ that were used in the ORA, are presented in S.Table 4.

We used R package *clusterProfiler* to perform the ORA analyses.[26] The settings/parameters that we used in ORA are shown in S.Table 5, and analysis-specific values of K and L are presented in pertinent table summarizing the results of the analysis.

Multiple testing corrections were applied on the generated *P*-values, for each exposure, separately, using the False Discovery Rate (FDR) approach, at the pathway level.[27] These were largely used to prioritize pathways in the ORA analysis, rather than the ADABF. By aggregating evidence adaptively across genes or variants, the Bayesian framework implemented in the ADABF approach inherently handles multiple testing.

To facilitate the presentation of the results, exposures were grouped by exposure category (S.Table 2) and the minimum *P*-value (across category sub-measurements) per gene/pathway was retained.

Further evidence of association with CRC risk for the genes included in the identified pathways was obtained from different sources. First, we used the gene ontology resource, and previously reported grouping scheme to assign enriched genes to hallmarks of cancer (Activating Invasion and Metastasis, Avoiding Immune Destruction, Deregulating Cellular Metabolism, Enabling Replicative Immortality, Evading Growth Suppressors, Genome Instability and Mutation, Resisting Cell Death, Sustaining Proliferative Signalling, Tumor-promoting Inflammation, Inducing or Accessing Vasculature).[28–30] Second, genes included within previously identified CRC risk loci were obtained from the latest GWASs.[31] Finally, Overall Association Scores were obtained from the Open Targets Platform (OTP). OTP is a publicly available resource that integrates a wide set of datasets, including genetic associations, somatic mutations, drugs, pathways and systems biology, RNA expression, text mining, and animal models.[32] Using this information, OTP can score genes with regard to an outcome on a scale of 0–1 (Overall Association Score), with values closer to one indicating stronger evidence. We used "Colorectal neoplasm" as the outcome term and a cut-off of 0.05 (corresponding to approximately the top 20%) to highlight genes with higher *a-priori* evidence.

All analyses were performed using R.[26,33]

Ethics

All participants gave written informed consent and studies were approved by their respective Institutional Review Boards.

Role of funders

The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

Results

Overview of associated and enriched pathways and genes

Associated pathways (ADABF)

Using the ADABF framework, a total of 1973 pathways were enriched ($P < 0.05$) out of the 2950 that were analysed for at least one exposure (S.Tables 6 and 7). Most enriched pathways were enriched for a single exposure: 136/181 (75.1%) of BIOCARTA, 96/139 (69.1%) of PID, 279/437 (63.8%) of WikiPathways, 667/1082 (61.6%) of REACTOME, and 67/134 (50%) of KEGG pathways. Several pathways were enriched for a combination of two exposures (21.0% of BIOCARTA to 39.6% of KEGG pathways), and few pathways were enriched for a combination of three (3.6% of PID to 10.4 of KEGG pathways) and four exposures (one WikiPathways and four REACTOME pathways). The majority were for combinations of BMI, smoking and NSAIDs (S.Table 8).

Enriched pathways (ORA)

Using ORA, 840 unique pathways, out of 2,950, were enriched ($P < 0.05$) across exposures and genes-sets, less than half of which for a single exposure: 20/37 (54.1%) of BIOCARTA, 37/88 (42.0%) of PID, 84/227 (37.0%) of WikiPathways, 154/385 (40.0%) of REACTOME, and 23/103 (22.3%) of KEGG pathways (S.Table 9).

Associated genes (ADABF)

Within the ADABF enriched pathways, a total of 1227 genes were associated ($P < 0.05$) with at least one exposure. Most of which were found for BMI (n = 768), followed by smoking (n = 223), NSAIDs (n = 173), fruit (n = 27), calcium (n = 25), red meat (n = 23), processed meat (n = 18), physical activity (n = 18), alcohol (n = 10), fibre (n = 8), T2D (n = 7), folate (n = 6), MHT (n = 5), and height (n = 3) (S.Table 10). Most of the genes (1142/1227; 93%) were associated with a single exposure, and 85 were associated with multiple exposures, the majority of which for combinations of BMI, smoking and NSAIDs (Fig. 2 and S.Table 11).

A total of 59 genes with a relatively lower a priori evidence for association with CRC (with an OTP ≤0.05 or not in the OTP database), were found to be enriched for any combination of exposures (Fig. 2 and S. Tables 10 and 11). Extensive overlaps, covering a wide range of genes [including cytochrome P450 family 4 subfamily B member 1 (*CYP4B1*), fibroblast growth factor 3 (*FGF3*), *GRIN2B*, integrin subunit alpha 6 (*ITGA6*), uncoupling protein 1 (*UCP1*), etc.] were found for BMI, and smoking. A similarly wide range of genes was enriched for BMI, and NSAIDs [such as inositol 1,4,5-trisphosphate receptor type 2 (*ITPR2*), protein kinase cGMP-dependent 1 (*PRKG1*), complement C3 (*C3*), claudin 11 (*CLDN11*), etc.]. Immune and stress-response related genes (such as Major Histocompatibility Complex coding genes, and members of the Heat Shock Protein family), as well as several genes of the Olfactory Receptor family [such as olfactory receptor family 1 subfamily N member 1 (*OR1N1*)] were commonly enriched for NSAIDs, and smoking. Multiple genes of the UDP glucuronosyltransferase family [such as UDP glucuronosyltransferase family 1 member A10 (*UGT1A10*)] involved in detoxification pathways were enriched for red meat intake and smoking. Among the overlaps were also genes associated with lipid metabolism, such as apolipoprotein E (*APOE*; enriched for red meat intake and smoking), acyl-CoA synthetase medium chain family member 3 (*ACSM3*; enriched for intakes of fibre and fruit), and low-density lipoprotein receptor (*LDLR*; enriched for MHT and red meat intake). Argonaute RISC component genes (such as *AGO1*, *AGO3*) involved in gene silencing were enriched for intakes of fruit and processed meat. Several taste receptor genes [such as taste 2 receptor member 20 (*TAS2R20*)], as well as gamma-aminobutyric acid type B receptor subunit 2 (*GABBR2*) were commonly enriched for BMI, and calcium intake. Several complement-related genes were enriched for fruit intake and smoking. Enrichment of *ARNTL* (basic helix-loop-helix ARNT like 1), implicated in circadian rhythms, was observed for NSAIDs and physical activity. Other enriched genes include: ferritin mitochondrial (*FTMT*; BMI, smoking, red meat), fibrinogen gamma chain (*FGG*; enriched for smoking, and processed meat intake), metallothionein 1M (*MT1M*; processed meat, physical activity), tight junction protein 1 (*TJP1*; BMI, fruit), tissue factor pathway inhibitor (*TFPI*; BMI, red meat), RB binding protein 8, endonuclease (*RBBP8*; BMI, alcohol), myocyte enhancer factor 2C (*MEF2C*; smoking, physical activity), and signal transducer and activator of transcription 5A (*STAT5A*; smoking and calcium).

The number of enriched genes within enriched pathways (ADABF) ranged from a minimum of one, to a maximum of 111, depending on the database, with a maximum of eight for BIOCARTA, 13 for PID, 37 for KEGG, 38 for WikiPathways, 111 for REACTOME (S.Table 12).

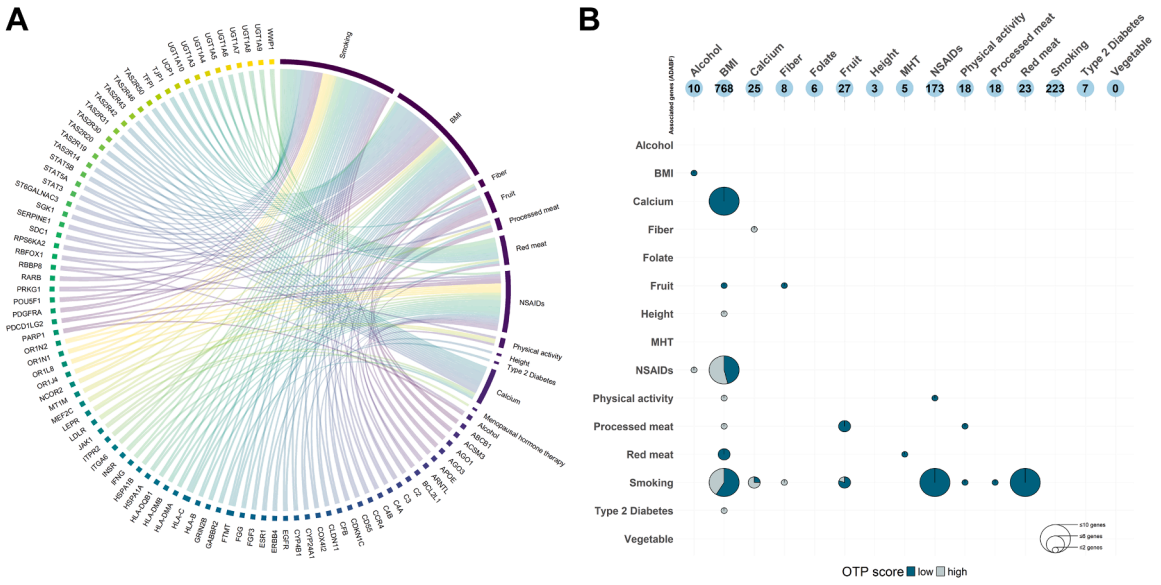


Fig. 2: Summary of enriched genes ($P < 0.05$) and genes commonly enriched across exposures using the adaptive combination of Bayes factors approach. **A.** Commonly enriched genes across different exposures. **B.** Total genes enriched per exposure (top), and number overlapping genes across exposures, by Open targets platform (OTP) score. BMI: Body mass index; NSAIDs: non-steroidal anti-inflammatory drugs. **List of gene abbreviations on the plot:** ABCB1: ATP binding cassette subfamily B member 1; ACSM3: acyl-CoA synthetase medium chain family member 3; AGO1: argonaute RISC component 1; AGO3: argonaute RISC catalytic component 3; APOE: apolipoprotein E; ARNTL: basic helix-loop-helix ARNT like 1; BCL2L1: BCL2 like 1; C2: complement C2; C3: complement C3; C4A: complement C4A (Chido/Rodgers blood group); C4B: complement C4B (Chido/Rodgers blood group); CCR4: C–C motif chemokine receptor 4; CD55: CD55 molecule (Cromer blood group); CDKN1C: cyclin dependent kinase inhibitor 1C; CFB: complement factor B; CLDN11: claudin 11; COX4I2: cytochrome *c* oxidase subunit 4I2; CYP24A1: cytochrome P450 family 24 subfamily A member 1; CYP4B1: cytochrome P450 family 4 subfamily B member 1; EGFR: epidermal growth factor receptor; ERBB4: erb-b2 receptor tyrosine kinase 4; ESR1: estrogen receptor 1; FGF3: fibroblast growth factor 3; FGG: fibrinogen gamma chain; FTMT: ferritin mitochondrial; GABBR2: gamma-aminobutyric acid type B receptor subunit 2; GRIN2B: glutamate ionotropic receptor NMDA type subunit 2B; HLA-B: major histocompatibility complex, class I, B; HLA-C: major histocompatibility complex, class I, C; HLA-DMA: major histocompatibility complex, class II, DM alpha; HLA-DMB: major histocompatibility complex, class II, DM beta; HLA-DQB1: major histocompatibility complex, class II, DQ beta 1; HSPA1A: heat shock protein family A (Hsp70) member 1A; HSPA1B: heat shock protein family A (Hsp70) member 1B; IFNG: interferon gamma; INSR: insulin receptor; ITGA6: integrin subunit alpha 6; ITPR2: inositol 1,4,5-trisphosphate receptor type 2; JAK1: Janus kinase 1; LDLR: low density lipoprotein receptor; LEPR: leptin receptor; MEF2C: myocyte enhancer factor 2C; MT1M: metallothionein 1M; NCOR2: nuclear receptor corepressor 2; OR1J4: olfactory receptor family 1 subfamily J member 4; OR1L8: olfactory receptor family 1 subfamily L member 8; OR1N1: olfactory receptor family 1 subfamily N member 1; OR1N2: olfactory receptor family 1 subfamily N member 2; PARP1: poly(ADP-ribose) polymerase 1; PDCD1LG2: programmed cell death 1 ligand 2; PDGFRA: platelet derived growth factor receptor alpha; POU5F1: POU class 5 homeobox 1; PRKG1: protein kinase cGMP-dependent 1; RARB: retinoic acid receptor beta; RBBP8: RB binding protein 8, endonuclease; RBFOX1: RNA binding fox-1 homolog 1; RPS6KA2: ribosomal protein S6 kinase A2; SDC1: syndecan 1; SERPINE1: serpin family E member 1; SGK1: serum/glucocorticoid regulated kinase 1; ST6GALNAC3: ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3; STAT3: signal transducer and activator of transcription 3; STAT5A: signal transducer and activator of transcription 5A; STAT5B: signal transducer and activator of transcription 5B; TAS2R14: taste 2 receptor member 14; TAS2R19: taste 2 receptor member 19; TAS2R20: taste 2 receptor member 20; TAS2R30: taste 2 receptor member 30; TAS2R31: taste 2 receptor member 31; TAS2R42: taste 2 receptor member 42; TAS2R43: taste 2 receptor member 43; TAS2R46: taste 2 receptor member 46; TAS2R50: taste 2 receptor member 50; TFPI: tissue factor pathway inhibitor; TJP1: tight junction protein 1; UCP1: uncoupling protein 1; UGT1A10: UDP glucuronosyltransferase family 1 member A10; UGT1A3: UDP glucuronosyltransferase family 1 member A3; UGT1A4: UDP glucuronosyltransferase family 1 member A4; UGT1A5: UDP glucuronosyltransferase family 1 member A5; UGT1A6: UDP glucuronosyltransferase family 1 member A6; UGT1A7: UDP glucuronosyltransferase family 1 member A7; UGT1A8: UDP glucuronosyltransferase family 1 member A8; UGT1A9: UDP glucuronosyltransferase family 1 member A9; WWP1: WW domain containing E3 ubiquitin protein ligase 1.

Highlighted pathways and genes

Top pathways

The top pathways based on the total number of enriched genes within the pathway, using the ADABF approach, are shown in Fig. 3. Seven hundred and fifty-two unique pathways had an FDR<5%, all across BMI (n = 688), NSAIDs (n = 68), and smoking (n = 53), the majority of which (698/752; 92.8%) for a single exposure (S.Table 7).

Using ORA, 152 unique pathways had an FDR<5% (S.Table 9). These pathways belonged to the libraries of KEGG (54/152; 35.5%), REACTOME (80/152; 52.6%),

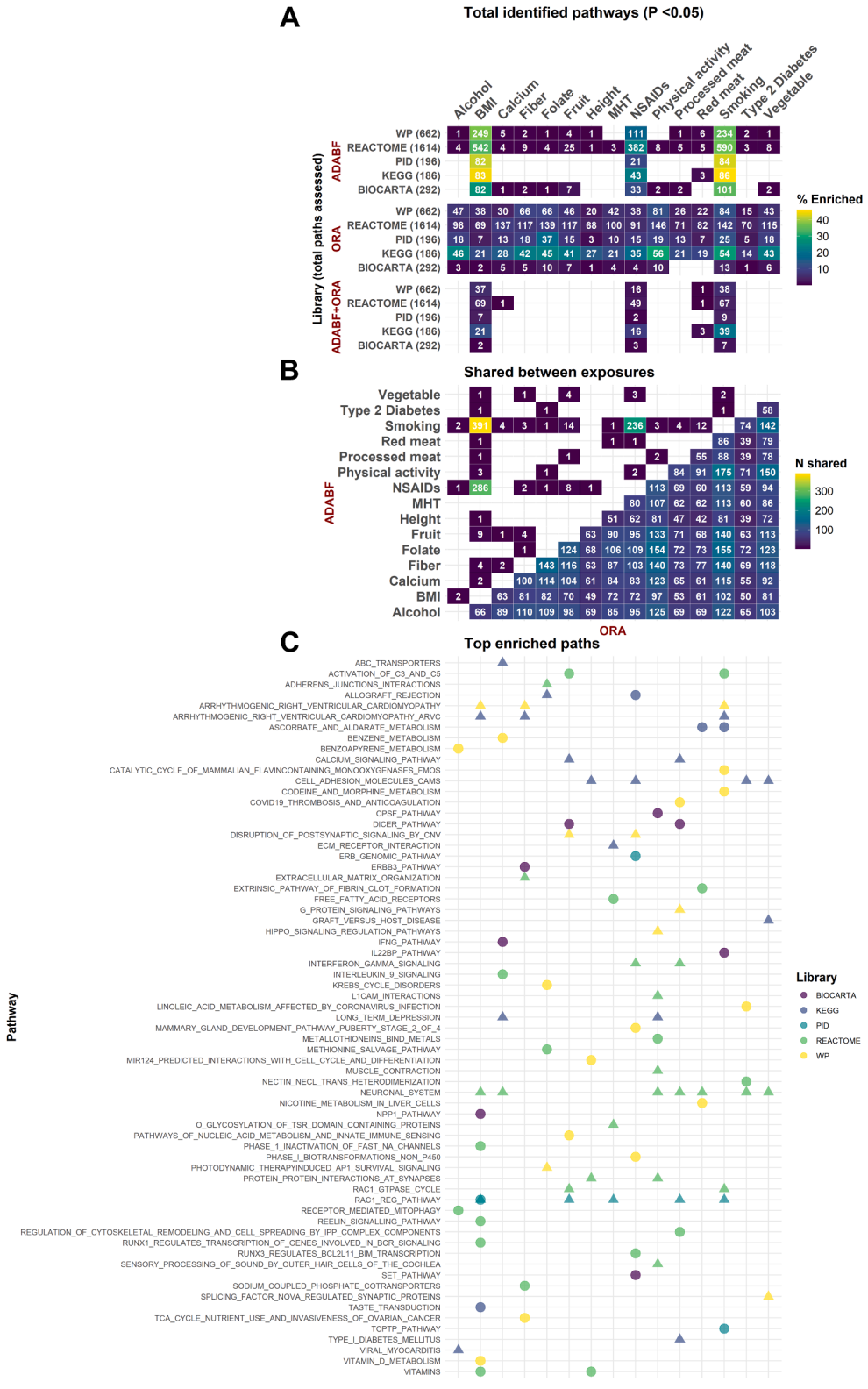


Fig. 3: Summary of enriched ($P < 0.05$) pathways. **A.** Total number of enriched pathways per gene-set library, exposure, and method used. Color is analogous to the percent of pathways enriched relative to pathways assessed. **B.** Overlapping of enriched pathways across exposures,

WikiPathways (16/152; 10.5%), and PID (2/152; 1.3%). Across gene-set libraries, 65 out of the 152 enriched pathways (42.8%) were below the 5% FDR threshold for a single exposure: 16/54 (29.6%) of KEGG, 40/80 (50.0%) of REACTOME, and 9/16 (56.3%) of WikiPathways. Thirty two of the 152 pathways (21.1%) had FDR < 5% for any combination of two exposures, and one pathway (REACTOME's *Neuronal system*) passed the 5% threshold for 14 different exposures (BMI, smoking, calcium, T2D, fibre, folate, fruit, height, MHT, NSAIDs, PA, processed meat, red meat, vegetables) (S.Table 9). Top pathways based on the minimum q-value are shown in Fig. 3.

A total of 297 pathways across five exposures (BMI, NSAIDs, smoking, calcium, and red meat) were enriched ($P < 0.05$) using both the ADABF and ORA approaches (S.Table 13). These pathways pertained to intra- and inter-cellular communication (such as the ErbB, Notch, MAPK, and epidermal growth factor (EGF) signalling pathways, and G-protein-mediated cascades), extracellular matrix organization, and cell adhesion [represented through pathways such as extracellular matrix organization, collagen biosynthesis and degradation, focal adhesion, and cell adhesion molecules (CAMs)], as well as regulation of the cytoskeleton (such as Integrin cell surface interactions). Immune Response and Inflammation (including interferon signalling, antigen processing and presentation), hormonal regulation pathways (such as aldosterone-regulated sodium reabsorption, insulin secretion and processing, and progesterone-mediated oocyte maturation), and pathways related to metabolism and biosynthetic processes (such as steroid hormone biosynthesis, glycosaminoglycan biosynthesis, and metabolism of steroids and fat-soluble vitamins), were also among the pathways that were identified using both ADABF and ORA approaches. Additionally, pathways involved in the nervous system (including those regulating neurotransmitter receptors and synaptic signalling, long-term potentiation and depression, and glutamate and GABA receptor activation), and pathways associated with cardiovascular and muscular functions, such as calcium signalling pathways, as well as hypertrophic and dilated cardiomyopathy pathways and cardiac conduction and contraction pathways were also included. Developmental pathways were also included, pertaining to processes such as nervous system development, neural crest cell migration, and pathways related to ectoderm and mesoderm development. Malignancy-associated pathways, such as pathways in cancer, small-cell and non-small-cell lung cancer, and melanoma, and other disease-specific pathways such as systemic lupus erythematosus and graft-versus-host disease, type I and type II diabetes mellitus, Alzheimer's disease, Rett syndrome were also in the list of commonly enriched pathways using the two approaches.

Top associated genes

Data were available for 811/1227 genes in the OTP; an overall association score >0.05 was found for 241 coding genes (S.Table 10). Most of the genes with score >0.05 were found for BMI (153/241), followed by smoking (47/241), NSAIDs (41/241), T2D (6/241), calcium (5/241, each), fruit (3/241), PA, alcohol, height, fibre, processed meat (2/241, each), and MHT (1/241).

Fifty percent of the genes (617/1227) mapped to at least one hallmark of cancer across schemes, most of which (388/617) pertained to the *Sustaining Proliferative Signalling* hallmark (S.Table 14). Exposure-wise, *Activating Invasion and Metastasis* was the top contributing hallmark for BMI (69% of genes), Avoiding Immune Destruction was the top contributing hallmark for smoking (28%) and NSAIDs (39%), and *Evading Growth Suppressors* was the top contributing hallmark for Calcium (5%), excluding categories with <5 participating genes.

Nine of the 1227 genes were previously identified as genome-wide significant CRC risk loci: DLG associated protein 1 (*DLGAP1*), pericentrin (*PCNT*), signal peptide, CUB domain and EGF like domain containing 1 (*SCUBE1*), vav guanine nucleotide exchange factor 2 (*VAV2*), Ras And Rab Interactor 3 (*RIN3*), Collagen Type IV Alpha 2 Chain (*COL4A2*), and Vesicle Transport Through Interaction With T-SNAREs 1A (*VTI1A*) enriched for BMI, Major Histocompatibility Complex, Class I, B (*HLA-B*) enriched for NSAIDs, Notch Receptor 4 (*NOTCH4*) enriched for fruit[15,31,34–36](S.Table 10).

Nearly half of the associated genes (689/1227; 56.2%) participated in pathways that were found to be enriched using both ADABF and ORA approaches, 177 of which (177/689; 25.7%) had an overall OTP score >0.05 (S.Table 13).

A selection of genes—and the pathways in which they are involved—including those not represented in the platform (to highlight potentially understudied associations) as well as genes with high OTP scores, is presented in Figs. 4 and 5.

Discussion

By utilizing large scale genome-wide gene-environment interaction estimates for 15 different exposures, we

per method used. **C.** Top enriched pathway per gene-set library, and exposure. Circles represent associated ADABF pathways with the size analogous to the percent of enriched genes within the pathway, and triangles represent enriched ORA pathways. The top ADABF pathways were defined by the total number of enriched genes within the pathway, while top ORA pathways based on the minimum q-value. ADABF: adaptive combination of Bayes factors; KEGG: Kyoto Encyclopedia of Genes and Genomes; ORA: Over-representation analysis; PID: Pathway Interaction Database; SNPs: Single nucleotide polymorphisms; WP: WikiPathways.

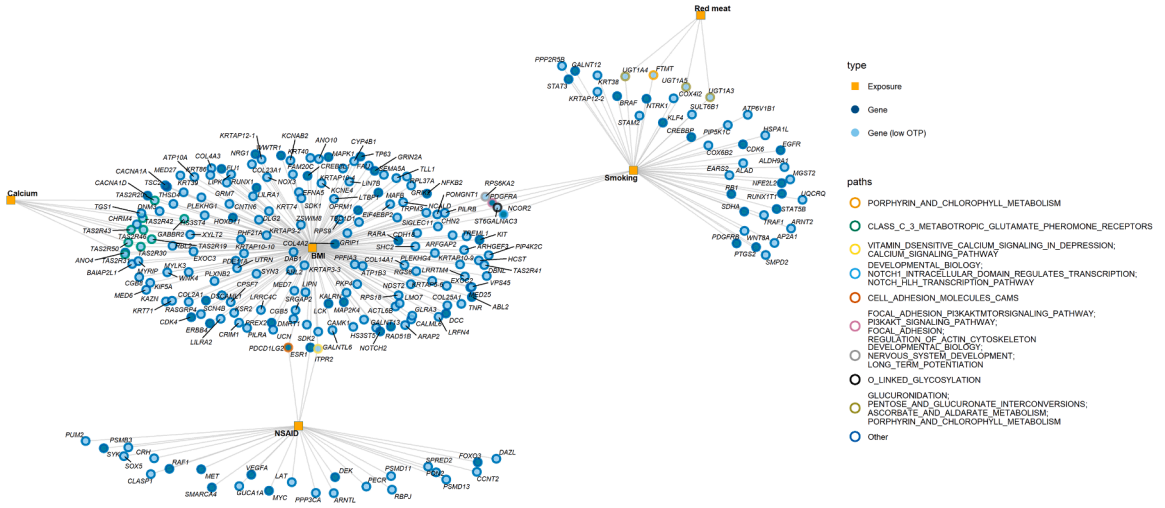


Fig. 4: Network of genes identified using the adaptive combination of Bayes factors (ADABF) within pathways that were highlighted by both the ADABF and over-representation analysis (ORA). The plot displays pathways containing genes commonly enriched across any combination of exposures. Human Leukocyte Antigen (HLA) genes were not included in the network. BMI: Body mass index; NSAIDs: non-steroidal anti-inflammatory drugs; OTP: open targets platform. **List of gene abbreviations on the plot:** ABL2: ABL proto-oncogene 2, non-receptor tyrosine kinase; ACTL6B: actin like 6B; ALAD: aminolevulinate dehydratase; ALDH9A1: aldehyde dehydrogenase 9 family member A1; ANO10: anoctamin 10; ANO4: anoctamin 4; AP2A1: adaptor related protein complex 2 subunit alpha 1; ARAP2: ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2; ARFGAP2: ADP ribosylation factor GTPase activating protein 2; ARHGEF3: Rho guanine nucleotide exchange factor 3; ARL2: ADP ribosylation factor like GTPase 2; ARNT2: aryl hydrocarbon receptor nuclear translocator 2; ATP10A: ATPase phospholipid transporting 10A (putative); ATP1B3: ATPase Na+/K+ transporting subunit beta 3; ATP6V1B1: ATPase H+ transporting V1 subunit B1; BAIAP2L1: BAR/IMD domain containing adaptor protein 2 like 1; ARNTL: basic helix-loop-helix ARNT like 1; BRAF: B-Raf proto-oncogene, serine/threonine kinase; CACNA1A: calcium voltage-gated channel subunit alpha1 A; CACNA1D: calcium voltage-gated channel subunit alpha1 D; CALML6: calmodulin like 6; CAMK1: calcium/calmodulin dependent protein kinase I; CCNT2: cyclin T2; CDH18: cadherin 18; CDK4: cyclin dependent kinase 4; CDK6: cyclin dependent kinase 6; CGB5: chorionic gonadotropin subunit beta 5; CGB8: chorionic gonadotropin subunit beta 8; CHN2: chimerin 2; CHRM4: cholinergic receptor muscarinic 4; CLASP1: cytoplasmic linker associated protein 1; CNTN6: contactin 6; COL14A1: collagen type XIV alpha 1 chain; COL23A1: collagen type XXIII alpha 1 chain; COL25A1: collagen type XXV alpha 1 chain; COL2A1: collagen type II alpha 1 chain; COL4A2: collagen type IV alpha 2 chain; COL4A3: collagen type IV alpha 3 chain; COX4I2: cytochrome *c* oxidase subunit 4I2; COX6B2: cytochrome *c* oxidase subunit 6B2; CPSF7: cleavage and polyadenylation specific factor 7; CREB3L1: cAMP responsive element binding protein 3 like 1; CREBBP: CREB binding protein; CRH: corticotropin releasing hormone; CRIM1: cysteine rich transmembrane BMP regulator 1; CYP4B1: cytochrome P450 family 4 subfamily B member 1; DAB1: DAB adaptor protein 1; DAZL: deleted in azoospermia like; DBNL: drebrin like; DCC: DCC netrin 1 receptor; DEK: DEK proto-oncogene; DLG2: discs large MAGUK scaffold protein 2; DMRT1: doublesex and mab-3 related transcription factor 1; DNM3: dynamin 3; DSCAML1: DS cell adhesion molecule like 1; EARS2: glutamyl-tRNA synthetase 2, mitochondrial; EFNA5: ephrin A5; EGFR: epidermal growth factor receptor; EIF4EBP2: eukaryotic translation initiation factor 4E binding protein 2; ERBB4: erb-b2 receptor tyrosine kinase 4; ESR1: estrogen receptor 1; EXOC2: exocyst complex component 2; EXOC3: exocyst complex component 3; FAM20C: FAM20C golgi associated secretory pathway kinase; FAU: FAU ubiquitin like and ribosomal protein S30 fusion; FLI1: Fli-1 proto-oncogene, ETS transcription factor; FOXO3: forkhead box O3; FTMT: ferritin mitochondrial; GABBR2: gamma-aminobutyric acid type B receptor subunit 2; GALNT12: polypeptide N-acetylgalactosaminyltransferase 12; GALNT13: polypeptide N-acetylgalactosaminyltransferase 13; GALNTL6: polypeptide N-acetylgalactosaminyltransferase like 6; GLRA3: glycine receptor alpha 3; GRIK2: glutamate ionotropic receptor kainate type subunit 2; GRIN2A: glutamate ionotropic receptor NMDA type subunit 2A; GRIP1: glutamate receptor interacting protein 1; GRM7: glutamate metabotropic receptor 7; GUCA1A: guanylate cyclase activator 1A; HCST: hematopoietic cell signal transducer; HOXD11: homeobox D11; HS3ST4: heparan sulfate-glucosamine 3-sulfotransferase 4; HS3ST5: heparan sulfate-glucosamine 3-sulfotransferase 5; HSPA1L: heat shock protein family A (Hsp70) member 1 like; ITPR2: inositol 1,4,5-trisphosphate receptor type 2; KALRN: kalirin RhoGEF kinase; KAZN: kazrin, periplakin interacting protein; KCNAB2: potassium voltage-gated channel subfamily A regulatory beta subunit 2; KCNE4: potassium voltage-gated channel subfamily E regulatory subunit 4; KIF5A: kinesin family member 5A; KIT: KIT proto-oncogene, receptor tyrosine kinase; KLF4: KLF transcription factor 4; KRT38: keratin 38; KRT39: keratin 39; KRT40: keratin 40; KRT71: keratin 71; KRT74: keratin 74; KRT86: keratin 86; KRTAP10-10: keratin associated protein 10-10; KRTAP10-4: keratin associated protein 10-4; KRTAP10-9: keratin associated protein 10-9; KRTAP12-1: keratin associated protein 12-1; KRTAP12-2: keratin associated protein 12-2; KRTAP3-2: keratin associated protein 3-2; KRTAP3-3: keratin associated protein 3-3; KRTAP5-6: keratin associated protein 5-6; KSR2: kinase suppressor of ras 2; LAT: linker for activation of T cells; LCK: LCK proto-oncogene, Src family tyrosine kinase; LILRA1: leukocyte immunoglobulin like receptor A1; LILRA2: leukocyte immunoglobulin like receptor A2; LIN7B: lin-7 homolog B, crumbs cell polarity complex component; LIPK: lipase family member K; LIPN: lipase family member N; LMO7: LIM domain 7; LRFN4: leucine rich repeat and fibronectin type III domain containing 4; LRRC4C: leucine rich repeat containing 4C; LRRTM4: leucine rich repeat transmembrane neuronal 4; LTBP1: latent transforming growth factor beta binding protein 1; MAFB: MAF bZIP transcription factor B;

agnostically investigated the molecular pathways underlying risk of CRC. Our analysis provides evidence to support the implication of several pathways and participating genes, few of which were previously identified as genome-wide significant CRC risk loci, potentially adding to the existing knowledge underlying the mechanisms behind CRC risk. Additionally, among the implicated genes and pertinent molecular pathways, concomitantly enriched pathways and genes for different exposures are highlighted, potentially suggesting common underlying mechanisms. The results of the present analysis provide a basis for further functional investigation. If confirmed, these findings could contribute to the development of personalized CRC prevention strategies. They could be used to inform refined risk stratification models that incorporate gene–environment interactions. Identifying genes and pathways whose effects are modifiable by lifestyle or environmental factors paves the way for targeted interventions aimed at mitigating genetic risk through behavioural or environmental modification.

Accumulating evidence over the past decades has established the importance of signalling pathways such as MAPK, Notch, PI3K/AKT pathway, TGF-β, and Wnt, in cell proliferation, differentiation, angiogenesis, invasion, apoptosis and survival in CRC.[37] Our analysis confirmed the above pathways and participating genes, several of which were overrepresented for exposures, such as smoking, BMI and NSAIDs. For instance, differential associations between smoking and CRC risk have been observed by pertinent categories of established molecular subtypes of CRC, such as mutations of B-Raf Proto-Oncogene, Serine/Threonine Kinase (*BRAF*) and KRAS Proto-Oncogene, GTPase (*KRAS*), CpG island methylator phenotype (CIMP), and microsatellite instability (MSI).[38] Genes closely linked to those molecular subtypes, such as *BRAF*, CREB Binding Protein (*CREBBP*), Epidermal Growth Factor Receptor (*EGFR*), Platelet Derived Growth Factor Receptor Beta (*PDGFRB*) and Neurotrophic Receptor Tyrosine Kinase 1 (*NTRK1*), were highlighted in interaction with smoking.[38–41] Additionally, genes that are known for

MAP2K4: mitogen-activated protein kinase kinase 4; MAPK1: mitogen-activated protein kinase 1; MED25: mediator complex subunit 25; MED27: mediator complex subunit 27; MED6: mediator complex subunit 6; MED7: mediator complex subunit 7; MET: MET proto-oncogene, receptor tyrosine kinase; MGST2: microsomal glutathione S-transferase 2; MYC: MYC proto-oncogene, bHLH transcription factor; MYLK3: myosin light chain kinase 3; MYRIP: myosin VIIA and Rab interacting protein; NCALD: neurocalcin delta; NCOR2: nuclear receptor corepressor 2; NDST2: N-deacetylase and N-sulfotransferase 2; NFE2L2: NFE2 like bZIP transcription factor 2; NFKB2: nuclear factor kappa B subunit 2; NOTCH2: notch receptor 2; NOX3: NADPH oxidase 3; NRG1: neuregulin 1; NTRK1: neurotrophic receptor tyrosine kinase 1; OPRM1: opioid receptor mu 1; PDCD1LG2: programmed cell death 1 ligand 2; PDE11A: phosphodiesterase 11A; PDGFRA: platelet derived growth factor receptor alpha; PDGFRB: platelet derived growth factor receptor beta; PECR: peroxisomal trans-2-enoyl-CoA reductase; PHF21A: PHD finger protein 21A; PILRA: paired immunoglobin like type 2 receptor alpha; PILRB: paired immunoglobin like type 2 receptor beta; PIP4K2C: phosphatidylinositol-5-phosphate 4-kinase type 2 gamma; PIP5K1C: phosphatidylinositol-4-phosphate 5-kinase type 1 gamma; PKP4: plakophilin 4; PLEKHG1: pleckstrin homology and RhoGEF domain containing G1; PLEKHG4: pleckstrin homology and RhoGEF domain containing G4; PLXNB2: plexin B2; POMGNT1: protein O-linked mannose N-acetylglucosaminyltransferase 1 (beta 1,2-); PON2: paraoxonase 2; PPFIA3: PTPRF interacting protein alpha 3; PPP2R5B: protein phosphatase 2 regulatory subunit B'beta; PPP3CA: protein phosphatase 3 catalytic subunit alpha; PREX2: phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 2; PSMB3: proteasome 20S subunit beta 3; PSMD11: proteasome 26S subunit, non-ATPase 11; PSMD13: proteasome 26S subunit, non-ATPase 13; PTGS2: prostaglandin-endoperoxide synthase 2; PUM2: pumilio RNA binding family member 2; RAD51B: RAD51 paralog B; RAF1: Raf-1 proto-oncogene, serine/threonine kinase; RARA: retinoic acid receptor alpha; RASGRP4: RAS guanyl releasing protein 4; RB1: RB transcriptional corepressor 1; RBL2: RB transcriptional corepressor like 2; RBPJ: recombination signal binding protein for immunoglobulin kappa J region; RGS6: regulator of G protein signalling 6; RPL37A: ribosomal protein L37a; RPS18: ribosomal protein S18; RPS6KA2: ribosomal protein S6 kinase A2; RPS9: ribosomal protein S9; RUNX1: RUNX family transcription factor 1; RUNX1T1: RUNX1 partner transcriptional co-repressor 1; SCN4B: sodium voltage-gated channel beta subunit 4; SDHA: succinate dehydrogenase complex flavoprotein subunit A; SDK1: sidekick cell adhesion molecule 1; SDK2: sidekick cell adhesion molecule 2; SEMA5A: semaphorin 5A; SHC2: SHC adaptor protein 2; SIGLEC11: sialic acid binding Ig like lectin 11; SMARCA4: SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4; SMPD2: sphingomyelin phosphodiesterase 2; SOX5: SRY-box transcription factor 5; SPRED2: sprouty related EVH1 domain containing 2; SRGAP2: SLIT-ROBO Rho GTPase activating protein 2; ST6GALNAC3: ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3; STAM2: signal transducing adaptor molecule 2; STAT3: signal transducer and activator of transcription 3; STAT5B: signal transducer and activator of transcription 5B; SULT6B1: sulfotransferase family 6B member 1; SYK: spleen associated tyrosine kinase; SYN3: synapsin III; TAS2R19: taste 2 receptor member 19; TAS2R20: taste 2 receptor member 20; TAS2R30: taste 2 receptor member 30; TAS2R31: taste 2 receptor member 31; TAS2R41: taste 2 receptor member 41; TAS2R42: taste 2 receptor member 42; TAS2R43: taste 2 receptor member 43; TAS2R46: taste 2 receptor member 46; TAS2R50: taste 2 receptor member 50; TBC1D1: TBC1 domain family member 1; TGS1: trimethylguanosine synthase 1; THSD4: thrombospondin type 1 domain containing 4; TLL1: tolloid like 1; TNR: tenascin R; TP63: tumour protein p63; TRAF1: TNF receptor associated factor 1; TREML1: triggering receptor expressed on myeloid cells like 1; TRPM3: transient receptor potential cation channel subfamily M member 3; TSC2: TSC complex subunit 2; UCN: urocortin; UGT1A3: UDP glucuronosyltransferase family 1 member A3; UGT1A4: UDP glucuronosyltransferase family 1 member A4; UGT1A5: UDP glucuronosyltransferase family 1 member A5; UQCRQ: ubiquinol-cytochrome c reductase complex III subunit VII; UTRN: utrophin; VEGFA: vascular endothelial growth factor A; VPS45: vacuolar protein sorting 45 homolog; WNK4: WNK lysine deficient protein kinase 4; WNT8A: Wnt family member 8A; WWTR1: WW domain containing transcription regulator 1; XYLT2: xylosyltransferase 2; ZSWIM8: zinc finger SWIM-type containing 8.

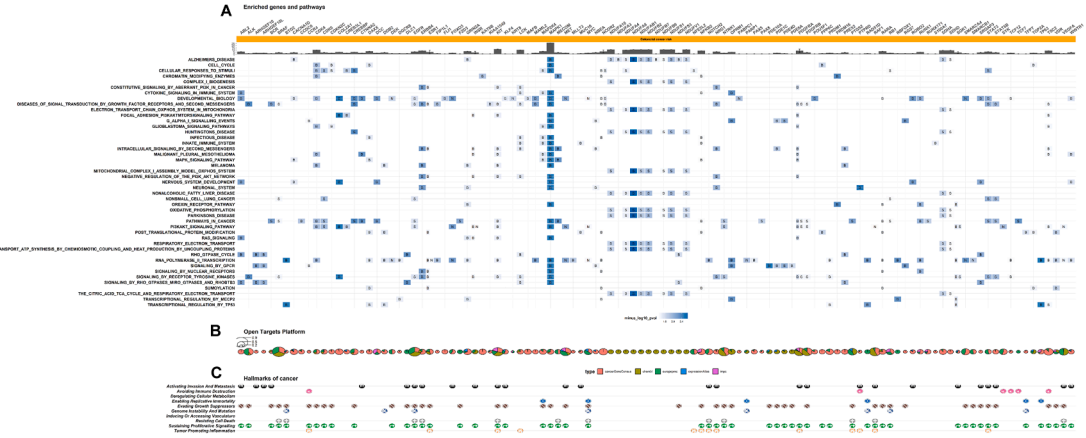


Fig. 5: Summary of the enriched genes and pathways using the adaptive combination of Bayes factors (ADABF) approach. Only the genes with an Open Targets Platform (OTP) overall association score >0.2, and pertinent pathways with >5 enriched genes with OTP>0.2 are plotted. **A**. Enriched genes and pathways. X-axis (top) presents enriched genes. Y-axis of the main panel shows the enriched pathways, and the tiles represent the exposures [B: Body mass Index (BMI); S: Smoking: N: Non-Steroidal Anti-inflammatory Drugs (NSAIDs)]. Tile color is analogous to the -log10 ADABF *P*-value of pertinent gene. **B**. Summary of the evidence from the OTP. Each pie-chart represents a gene and illustrates the relative contribution of individual components that form the overall OTP association score, and their size is analogous to the OTP overall association score. Five sub-components of OTP are plotted, cancerGeneCensus: representing genes containing mutations that have been causally implicated in cancer, chembl: representing bioactive molecules with drug-like properties, europepmc: representing gene-colorectal cancer (CRC) co-occurrences in the literature, expressionAtlas: representing differentially expressed genes in CRC, impc: representing the similarity between a mouse model and a human CRC on each gene. **C**. Summary of the Hallmarks of cancer. The bottom panel shows the hallmark capabilities in which each gene has been implicated in based on various grouping schemes. **List of gene abbreviations on the plot:** ABL2: ABL proto-oncogene 2, non-receptor tyrosine kinase; ALK: ALK receptor tyrosine kinase; ARHGEF10: Rho guanine nucleotide exchange factor 10; ARHGEF10L: Rho guanine nucleotide exchange factor 10 like; BCR: BCR activator of RhoGEF and GTPase; BRAF: B-Raf proto-oncogene, serine/threonine kinase; BTG2: BTG anti-proliferation factor 2; CACNA1D: calcium voltage-gated channel subunit alpha1 D; CCDC6: coiled-coil domain containing 6; CCR4: C–C motif chemokine receptor 4; CDK4: cyclin dependent kinase 4; CDK6: cyclin dependent kinase 6; CDKN2C: cyclin dependent kinase inhibitor 2C; COL2A1: collagen type II alpha 1 chain; CREB3L1: cAMP responsive element binding protein 3 like 1; CREBBP: CREB binding protein; CTNNA2: catenin alpha 2; DAXX: death domain associated protein; DCC: DCC netrin 1 receptor; DDB2: damage specific DNA binding protein 2; DEK: DEK proto-oncogene; DGCR8: DGCR8 microprocessor complex subunit; DLC1: DLC1 Rho GTPase activating protein; EGFR: epidermal growth factor receptor; ERBB4: erb-b2 receptor tyrosine kinase 4; ESR1: estrogen receptor 1; FIP1L1: factor interacting with PAPOLA and CPSF1; FLI1: Fli-1 proto-oncogene, ETS transcription factor; FOXO3: forkhead box O3; FZD3: frizzled class receptor 3; GRIN2A: glutamate ionotropic receptor NMDA type subunit 2A; INSR: insulin receptor; KAT6B: lysine acetyltransferase 6B; KIAA1549: KIAA1549; KIT: KIT proto-oncogene, receptor tyrosine kinase; KLF4: KLF transcription factor 4; KRT8: keratin 8; LCK: LCK proto-oncogene, Src family tyrosine kinase; MAFB: MAF bZIP transcription factor B; MAML2: mastermind like transcriptional coactivator 2; MAP2K4: mitogen-activated protein kinase kinase 4; MAPK1: mitogen-activated protein kinase 1; MECOM: MDS1 and EVI1 complex locus; MET: MET proto-oncogene, receptor tyrosine kinase; MLLT3: MLLT3 super elongation complex subunit; MUC16: mucin 16, cell surface associated; MYC: MYC proto-oncogene, bHLH transcription factor; NBEA: neurobeachin; NCOR2: nuclear receptor corepressor 2; NDUFA10: NADH:ubiquinone oxidoreductase subunit A10; NDUFA3: NADH:ubiquinone oxidoreductase subunit A3; NDUFA4: NDUFA4 mitochondrial complex associated; NDUFA8: NADH:ubiquinone oxidoreductase subunit A8; NDUFAB1: NADH:ubiquinone oxidoreductase subunit AB1; NDUFB2: NADH:ubiquinone oxidoreductase subunit B2; NDUFB7: NADH:ubiquinone oxidoreductase subunit B7; NDUFB9: NADH:ubiquinone oxidoreductase subunit B9; NDUFS1: NADH:ubiquinone oxidoreductase core subunit S1; NDUFS3: NADH:ubiquinone oxidoreductase core subunit S3; NDUFV1: NADH:ubiquinone oxidoreductase core subunit V1; NFE2L2: NFE2 like bZIP transcription factor 2; NFKB2: nuclear factor kappa B subunit 2; NOTCH2: notch receptor 2; NRG1: neuregulin 1; NTRK1: neurotrophic receptor tyrosine kinase 1; OPRM1: opioid receptor mu 1; PABPC1: poly(A) binding protein cytoplasmic 1; PARP1: poly(ADP-ribose) polymerase 1; PAX5: paired box 5; PAX8: paired box 8; PDE10A: phosphodiesterase 10A; PDE1C: phosphodiesterase 1C; PDE4D: phosphodiesterase 4D; PDE8A: phosphodiesterase 8A; PDGFRA: platelet derived growth factor receptor alpha; PDGFRB: platelet derived growth factor receptor beta; POU5F1: POU class 5 homeobox 1; PPP6C: protein phosphatase 6 catalytic subunit; PRDM1: PR/SET domain 1; PRDM16: PR/SET domain 16; PREX2: phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 2; PTGS2: prostaglandin-endoperoxide synthase 2; PTPRD: protein tyrosine phosphatase receptor type D; RAD51D: RAD51 paralog D; RAF1: Raf-1 proto-oncogene, serine/threonine kinase; RARA: retinoic acid receptor alpha; RB1: RB transcriptional corepressor 1; RBFOX1: RNA binding fox-1 homolog 1; RGS7: regulator of G protein signalling 7; ROBO2: roundabout guidance receptor 2; RUNX1: RUNX family transcription factor 1; RUNX1T1: RUNX1 partner transcriptional co-repressor 1; S100A7: S100 calcium binding protein A7; SDHA: succinate dehydrogenase complex flavoprotein subunit A; SDHD: succinate dehydrogenase complex subunit D; SGK1: serum/glucocorticoid regulated kinase 1; SMARCA4: SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4; SMARCB1: SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1; SRGAP3: SLIT-ROBO Rho GTPase activating protein 3; STAT3: signal transducer and activator of transcription 3; STAT5B: signal transducer and activator of transcription 5B; SYK: spleen associated tyrosine kinase; TCF12: transcription factor 12; TCF7: transcription factor 7; TFPT: TCF3 fusion partner; TOP2A: DNA topoisomerase II alpha; TP63: tumour protein p63; TSC2: TSC complex subunit 2; VDR: vitamin D receptor; VEGFA: vascular endothelial growth factor A; WWTR1: WW domain containing transcription regulator 1.

their crucial role in carcinogenesis, such as the oncogenes Calcium Voltage-Gated Channel Subunit Alpha1 D (*CACNA1D*), KIT Proto-Oncogene, Receptor Tyrosine Kinase (*KIT*), Mitogen-Activated Protein Kinase Kinase 4 (*MAP2K4*) and *COL4A2*, emerged in interaction with BMI, but their implication in the BMI to CRC risk association are yet to be elucidated.[42–45] Important modulators of the anti-tumour immune response, such as Spleen Associated Tyrosine Kinase (*SYK*) and Vascular Endothelial Growth Factor A (*VEGFA*), and several other inflammatory and pro-inflammatory cytokine coding genes [such as interleukin (IL)-9, *IL12B,* CD40 Molecule (*CD40*), *VEGFA,* TNF Superfamily Member 10 (*TNFSF10*)], and related immune-regulatory paths, were highlighted for NSAIDs.[46] NSAIDs can inhibit the activation of cancer pathways (NF-κB) related to increased proliferation and survival, independently of COX-2 inhibition (which is thought to be the primary mechanism via which NSAIDs exert their anti-inflammatory and anti-tumour effects).[47–49] In addition, several oncogenes like MET Proto-Oncogene, Receptor Tyrosine Kinase (*MET*), Raf-1 Proto-Oncogene, Serine/Threonine Kinase (*RAF1*), MYC proto-oncogene, bHLH transcription factor (*MYC*) were enriched for NSAIDs.[50,51] Many of the above genes and their gene products have been previously identified as potential therapeutic drug targets of high priority, and the results of the present analysis provide additional evidence to support their association with CRC risk.[52] Moreover, our analysis identified several other genes that are less studied in relation to the investigated exposures, such as argonaute RISC component, UDP-glucuronosyltransferase, and taste receptor coding genes, however their role in CRC tumorigenesis remains to be explained.

Overall, the results of the present analysis are characteristic of the breadth of mechanisms via which modifiable exposures, like excess body weight and smoking, act to increase CRC risk.[53] Several key molecular pathways, that fall under the *sustaining proliferative signalling* and *evading growth suppressors* hallmark capabilities, such as the PI3K-Akt and MAPK signalling pathway, signalling by GPCR and Receptor Tyrosine Kinases (RTK), were enriched in the analysis for BMI and smoking.[54,55] Pre-clinical models have shown that increased adiposity may drive epithelial mesenchymal transition (EMT), while various groups of active compounds in cigarette smoke can induce EMT via different signalling pathways.[56,57] Several genes [Notch Receptor 2 (*NOTCH2*), mitogen-activated protein kinase 1 (*MAPK1*), Collagen Type IV Alpha 1 Chain (*COL4A1*)] and pathways that are involved in EMT, such as the Notch, TGF-beta, Wnt–β-catenin, Hedgehog, RTK, were also enriched for BMI and smoking.[56] Obesity has been associated with cell structure changes that render the tissue microenvironment more favourable for tumour growth and metastasis.[58] Genes such as Collagen Type IV Alpha 3 Chain (*COL4A3*), platelet derived growth factor receptor alpha (*PDGFRA*), syndecan 1 (*SDC1*), and pertinent pathways potentially implicated in the process of tissue remodelling were enriched for BMI.[59] Additionally, increased adiposity is a condition characterized by metabolic dysfunction that results in *deregulation of cellular energetics*, another hallmark of cancer.[60] Obesity affects the secretion of various endogenous hormones, like insulin and Insulin Like Growth Factor 1 (*IGF1*), estrogens, leptin and other adipokines, that apart from the effect they have on downstream proliferation-promoting cascades, also affect metabolism-related pathways.[61] Several pathways directly related to metabolic reactions (metabolism of lipids, carbohydrates, glycolysis and gluconeogenesis), as well as hormone receptor coding genes, such as the Leptin Receptor (*LEPR*), insulin receptor (*INSR*), estrogen receptor 1 (*ESR1*), and hormone-associated metabolic pathways were enriched for BMI.

Concomitantly enriched pathways and genes for different exposures were highlighted in our study. There was a notable overlap of enriched genes and pathways between BMI and NSAIDs, which suggests that potentially the BMI associated pathways may be particularly relevant to the aspirin/NSAID related protective association. Apart from key cancer-associated genes and pathways, such as BCL2 like 1 (*BCL2L1*; via cytokine-mediated signalling), and the Jak/STAT and PI3K-Akt pathways, several immune-regulatory, hormone-related, as well as cell-structure and motility-associated pathways were enriched. Adipose tissue contains numerous immune cells that create an inflammatory microenvironment, and several experimental studies have demonstrated a link between obesity and immunoregulatory proteins [such as programmed cell death 1 ligand 2 (*PDCD1LG2*)].[62] Such adipose tissue derived inflammation contributes to insulin resistance, while inflammatory proteins may interact with adipokines and estrogens.[63–65] Interestingly, hormone receptors like *INSR, LEPR, ESR1* were concomitantly enriched for BMI and NSAIDs. *SDC1,* a key cell surface adhesion molecule, that plays an important role in maintaining cell morphology, was also commonly enriched for those two exposures.[66] Previous observational studies have shown a potential interaction of smoking and BMI in the NSAID-CRC association.[67,68] The protective effect of aspirin on CRC was stronger among never-smokers compared to smokers, and the association between any NSAID use and CRC risk was attenuated with increasing BMI.[68–70] Bitter taste receptor coding genes, such as Taste 2 Receptor Member 14 (*TAS2R14*), and Taste 2 Receptor Member 46 (*TAS2R46*) were highlighted in interaction with both BMI and calcium. Emerging evidence suggests that bitter taste receptors may play a role in modulating early tumorigenic processes in colorectal

cancer.[71] Potentially important commonalities observed for BMI and smoking include nuclear receptor corepressor 2 (*NCOR2*), a gene that encodes for a nuclear receptor co-repressor, which mediates transcriptional silencing of various genes, and *PDGFRA*, a gene that encodes for a cell surface tyrosine kinase receptor of platelet-derived growth factors.[72,73] Both these genes have been characterized as promising therapeutic targets for CRC as they have received high OTP overall association scores (>0.5). Recent studies have linked both obesity and smoking with epigenetic alterations, and differential mRNA expression of *NCOR2,* but evidence for a common mechanism for *PDGFRA* is limited.[74,75] Few commonalities were observed between BMI and physical activity. Notable ones include BIOCARTA's circadian pathway, identified using the ADABF approach, as well as pathways related to extracellular matrix organization, identified using ORA. An interesting commonality for NSAIDs and physical activity via the Basic Helix-Loop-Helix ARNT Like 1 (*BMAL1/ARNTL1)* gene, a transcriptional activator forming a core component of the circadian clock, was also observed. Regular physical activity positively influences circadian rhythmicity, and NSAID use has also been shown to affect the expression of circadian clock genes, whereas circadian disruption has been associated with increased risk of CRC.[76–78] STAT protein coding genes [signal transducer and activator of transcription 3 (*STAT3*), signal transducer and activator of transcription 5B (*STAT5B*)], and related pathways, such as the IL9 and IL21 signalling pathways, that mediate cellular responses to interferons, and likely involved in the *Sustained proliferative signalling* hallmark, were enriched for calcium and smoking. In a potentially similar context, IFN gamma signalling pathway [and the Janus kinase 1 (*JAK1*) gene], was enriched for both calcium and fibre. Poly(ADP-ribose) polymerase 1 (*PARP1*), a gene that mediates significant post-translational modification of proteins that plays a key role in DNA repair, falling under the *Enabling Replicative Immortality*, and *Genome Instability and Mutation* hallmarks was enriched for both Alcohol and NSAIDs, potentially suggesting a common mechanism between the two exposures. We found a processed meat with fruit intake commonality, of the Argonaute protein coding genes Argonaute RISC Component 1 (*AGO1*) and Argonaute RISC Component 3 (*AGO3*), within the Dicer pathway, that are used for RNA-mediated gene silencing. Several genes from the UDP-glucuronosyltransferase family, including UDP Glucuronosyltransferase Family 1 Member A4 (*UGT1A4*), were identified in relation to both red meat intake and smoking, potentially indicating shared mechanisms involving xenobiotic metabolism.[79] However, there is limited evidence in the current literature to support these as common underlying pathways for CRC risk. Further research is needed to determine whether the observed overlaps reflect true shared biological mechanisms. Gaining a deeper understanding of these potential commonalities could help refine cancer prevention guidelines and support the development of more targeted preventive strategies, particularly for individuals at higher CRC risk.

Although we used a relatively relaxed threshold for SNPs mapping to genes as being significant in the ORA analysis, there still might be significant loss of information due to the imposed *P*-value cutoff, and genes with marginally significant signals may have been ignored. This is a known limitation of ORA. On the other hand, the ADABF incorporates continuous measures of evidence strength (betas and corresponding variances from the interaction models), which is an advantage of association analysis. It is also well known that a method testing the association null hypothesis has a larger power compared to ORA which is an enrichment analysis method. This means that the ADABF method is more likely to reject the null hypothesis for more pathways than an enrichment method. These method characteristics can explain the differences in the results between the ORA and ADABF approaches, and the fact that more genes and pathways were found to be associated with exposures with stronger effects with CRC risk, like BMI, and smoking. Although pathways identified by both methods can be considered robustly enriched and prioritized for CRC risk, those identified by only one method should not be disregarded. Instead, they should be explored complementarily in future investigations. Even though we restricted our analysis to individuals of European ancestry and adjusted for three principal components to capture remaining variability, subtle population structure may still confound GWAS results, particularly for traits influenced by environmental or lifestyle factors. Such residual stratification could bias genetic associations and potentially lead to spurious findings.

The large sample size, and the high-quality phenotype and genotype data may minimize potential sources of bias and are among the strengths of our study. A wide range of established CRC risk factors were included, and several gene-set libraries were used to cover a variety of biological pathways, in the pathway enrichment analysis. However, exposure assessments were self-reported in most of the studies included, measured at a single time point, and may have been subject to measurement errors. Additionally, the estimates on which the analysis was based were derived exclusively from European ancestry populations, which limits generalizability. To maximize statistical power, our analysis focused on CRC overall, without performing stratified analyses by subsite (e.g., right colon, left colon, and rectum), sex, or early-onset CRC. Given that associations with CRC risk may vary across these subgroups and considering the shifting prevalence of CRC risk factors in populations over time, future

studies should explore these associations within specific strata.

The results of the present analysis may aid in elucidating the etiological associations of risk factors to CRC risk and inform personalized prevention strategies.

Contributors

EB: Data curation, Formal analysis, Methodology, Investigation, Visualization, Writing–Original Draft, Writing–Review & Editing; RY: Data verification, Data curation, Formal analysis, Writing–Review & Editing; AEK: Data verification, Data curation, Formal analysis, Writing–Review & Editing; GM, NM, DA, LNA, ELB, SIB, TDB, HB, ABH, PTC, RCT, ATC, IC, MAD, VDO, ND, DAD, SBG, AG, MH, LH, JRH, EK, TOK, AK, SK, LLM, JPL, LL, BML, VM, JM, JM, CCN, MOS, JRP, NP, AJP, ARP, PDPP, EAP, CQ, ERN, JSM, RES, MCS, CET, YT, CYM, KV, PV, VV, EW, AW, MOW, AHW, MJG: Methodology, Writing–Review & Editing; WJG, UP, ME: Conceptualization, Investigation, Methodology, Writing–Review & Editing, Supervision; KKT: Conceptualization, Funding acquisition, Investigation, Methodology, Writing–Review & Editing, Supervision. All authors critically revised the manuscript for intellectual content. All authors read and approved the final manuscript.

Data sharing statement

Data described in the manuscript will be made available upon request pending application and approval to the GECCO consortium.

Declaration of interests

EB was supported by grants from Cancer Research UK (CRUK; grants number: PPRCPJT\100005), and World Cancer Research Fund International (WCRF; IIG_FULL_2020_022). RY was supported by Cancer Research UK (CRUK; grants number: PPRCPJT\100005). ELB was supported by the National Cancer Institute, National Institutes of Health. SB was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH. ATC was supported by grants from the National Institutes of Health, Stand Up to Cancer, Freenome Holdings, and American Cancer Society; Received payment for expert testimony by Boehringer Ingelheim; Participation on a Data Safety Monitoring Board or Advisory Board for Pfizer Inc; Held leadership or fiduciary role in the American Gastroenterological Association, and National Cancer Institute. DAD was supported by a National Institutes of Health grant. ESK was supported by grants from the National Cancer Institute [R01 CA201407, P01 CA196569]; Holds stocks or stock options from Abbvie (ABBV), and Pfizer (PFE). TOK was supported by a National Cancer Institute grant [R01CA066635, 1996–2008]. AK was supported by a National Institutes of Health grant [R01CA273198]; Was a consultant with Bristol Myers Squibb, Inari Agriculture, and Arcadia Science; Holds stocks or stock options from TensorBio Inc, SerImmune Inc., Illumina, Deep Genomics, ImmunAI. BML was a consultant with Elsevier; Was supported for attending meetings and/or travel by the International Epidemiological Association; Held leadership or fiduciary role in the International Epidemiological Association Council. VM was supported by grants from the Instituto de Salud Carlos III, Programa FORTALECE del Ministerio de Ciencia e Innovación through the project number FORT23/00032, Spanish Association Against Cancer (AECC) Scientific Foundation grant GCTRA18022MORE, and Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP) action Genrisk. JLM was supported by grants from the National Cancer Institute [R01 CA201407, P01 CA196569, R01 CA273198]. JRP was supported by a National Institutes of Health grant. AJP Was a consultant with Abbvie. PDPP was supported by a Cancer Research UK grant. EAP held leadership or fiduciary role (unpaid), as an elected member of the American Association for Cancer Research Board of Directors. RES was supported by grants from the National Cancer Institute NCI EDRN Grant (U01-CA152753), and Exact Sciences, Freenome, Immunovia; Was a consultant with Guardant, Freenome, Exact Sciences; Participated on a Data Safety Monitoring Board or Advisory Board in VA CONFIRM trial, NCI Cancer Screening Research Network. CET is an epidemiology contractor with Pfizer, unrelated to the present study. VV was supported by a Czech Science Foundation 21-04607X grant, unrelated to the submitted work. AW was supported by a Swedish Cancer Foundation grant. WJG was supported by grants from the National Cancer Institute [R01 CA201407, P01 CA196569, R01 CA273198]. UP was supported by grants from the National Institute of Health, V Foundation, and Goldman Sachs Foundation; Was a consultant with AbbVie; UP's family is holding individual stocks for the following companies: Amazon, ARM Holdings PLC, BioNTech, BYD Company Limited, Crowdstrike Holdings Inc, CureVac, Google/Alphabet, Microsoft Corp, NVIDIA Corp, Stellantis. ME was supported by Cancer Research UK (CRUK; grants number: PPRCPJT\100005). KKT was supported by grants from Cancer Research UK (CRUK; grants number: PPRCPJT\100005), and World Cancer Research Fund International (WCRF; IIG_FULL_2020_022).The remaining authors have no competing interests to declare.

Acknowledgements

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. This article is the result of the scientific work of Dr Murphy while he was affiliated at IARC.

The authors assume full responsibility for all analyses and interpretation of results. The views expressed here are those of the authors and do not necessarily represent the American Cancer Society or the American Cancer Society—Cancer Action Network.

This research was supported [in part] by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH author(s) were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088, R01 CA059045, U01 CA164930, R21 CA191312, R01 CA244588, R01 CA206279, R01 CA201407, R01 CA488857, P20 CA252733). Genotyping/Sequencing services were provided by the Center for Inherited Disease Research (CIDR) contract number HHSN268201700006I and HHSN268201200008I. This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704. Scientific Computing Infrastructure at Fred Hutch funded by ORIP grant S10OD028685.

ASTERISK: a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC).

The ATBC Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

CLUE II funding was from the National Cancer Institute (U01 CA086308, Early Detection Research Network; P30 CA006973), National Institute on Aging (U01 AG018033), and the American Institute for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. Maryland Cancer Registry (MCR) Cancer data were provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The

collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

The Colon Cancer Family Registry (CCFR, www.coloncfr.org) is supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). Support for case ascertainment was provided in part from the Surveillance, Epidemiology, and End Results (SEER) Program and the following U.S. state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada). The CCFR Set-1 (Illumina 1M/1M-Duo) was supported by NIH awards U01 CA122839 and R01 CA143237 (to GC). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and R01 CA81488 (to SBG). The CCFR Set-4 (Illumina OncoArray 600K SNP array) was supported by NIH award U19 CA148107 (to SBG) and by the Center for Inherited Disease Research (CIDR), which is funded by the NIH to the Johns Hopkins University, contract number HHSN2682012000081. The content of this manuscript does not necessarily reflect the views or policies of the NCI, NIH or any of the collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the CCFR.

COLO2&3: National Institutes of Health (R01 CA060987).

Colorectal Cancer Transdisciplinary (CORECT) Study: The CORECT Study was supported by the National Cancer Institute, National Institutes of Health (NCI/NIH), U.S. Department of Health and Human Services (grant numbers U19 CA148107, R01 CA081488, P30 CA014089, R01 CA197350; P01 CA196569; R01 CA201407; R01 CA242218), National Institute of Environmental Health Sciences, National Institutes of Health (grant number T32 ES013678) and a generous gift from Daniel and Maryann Fong.

CORSA: The CORSA study was funded by Austrian Research Funding Agency (FFG) BRIDGE (grant 829675, to Andrea Gsur), the "Herzfelder'sche Familienstiftung" (grant to Andrea Gsur) and was supported by COST Action BM1206.

CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (CPS-II) cohort. The study protocol was approved by the institutional review boards of Emory University, and those of participating registries as required.

CRCGEN: Colorectal Cancer Genetics & Genomics, Spanish study was supported by Instituto de Salud Carlos III, co-funded by FEDER funds—a way to build Europe—(grants PI14-613 and PI09-1286), Agency for Management of University and Research Grants (AGAUR) of the Catalan Government (grant 2017SGR723), Junta de Castilla y León (grant LE22A10-2), the Spanish Association Against Cancer (AECC) Scientific Foundation grant GCTRA18022MORE and the Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), action Genrisk. Sample collection of this work was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d'Oncología de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology. We thank CERCA Programme, Generalitat de Catalunya for institutional support.

Czech Republic CCS: This work was supported by the Czech Science Foundation (21-04607X, 21-27902S), by the Grant Agency of the Ministry of Health of the Czech Republic (grants AZV NU21-07-00247 and AZV NU21-03-00145), and Charles University Research Fund (Cooperation 43-Surgical disciplines).

DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1 and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B).

DALS: National Institutes of Health (R01 CA048998 to M. L. Slattery).

EDRN: This work is funded and supported by the NCI, EDRN Grant (U01-CA152753).

EPIC: The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and also 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 Region Skåne and Region 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). (United Kingdom).

ESTHER/VERDI. This work was supported by grants from the Baden-Württemberg Ministry of Science, Research and Arts and the German Cancer Aid.

Harvard cohorts: HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), NHS by the National Institutes of Health (P01 CA087969, UM1 CA186107, R01 CA137178, R01 CA151993, and R35 CA197735), and PHS by the National Institutes of Health (R01 CA042182).

Kentucky: This work was supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726.

LCCS: The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167).

MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 509348, 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. BMLynch was supported by MCRF18005 from the Victorian Cancer Agency.

MEC: National Institutes of Health (R37 CA054281, P01 CA033619, and R01 CA063464).

MECC: This work was supported by the National Institutes of Health, U.S. Department of Health and Human Services (R01 CA081488, R01 CA197350, U19 CA148107, R01 CA242218, and a generous gift from Daniel and Maryann Fong.

NCCCS I & II: We acknowledge funding support for this project from the National Institutes of Health, R01 CA066635 and P30 DK034987.

NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821); the National Institutes of Health, U.S. Department of Health and Human Serivces (U01 CA074783); and National Cancer Institute of Canada grants (18223 and 18226). The authors wish to acknowledge the contribution of Alexandre Belisle and the genotyping team of the McGill University and Génome Québec Innovation Centre,

Montréal, Canada, for genotyping the Sequenom panel in the NFCCR samples. Funding was provided to Michael O. Woods by the Canadian Cancer Society Research Institute.

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

PPS3 and PPS4 Polyp Prevention Study trials: were funded by the National Cancer Institute, National Institutes of Health (grant numbers R01CA059005 and R01CA098286).

SEARCH: The University of Cambridge has received salary support in respect of PDPP from the NHS in the East of England through the Clinical Academic Reserve. Cancer Research UK (C490/A16561); the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge.

SELECT: Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Numbers U10 CA037429 (CD Blanke), and UM1 CA182883 (CM Tangen/IM Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

SMS and REACH: This work was supported by the National Cancer Institute (grant P01 CA074184 to J.D.P. and P.A.N., grants R01 CA097325, R03 CA153323, and K05 CA152715 to P.A.N., and the National Center for Advancing Translational Sciences at the National Institutes of Health (grant KL2 TR000421 to A.N.B.-H.)

Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the Swedish Research Council/Infrastructure grant, the Swedish Cancer Foundation, and the Karolinska Institute's Distinguished Professor Award to Alicja Wolk.

UK Biobank: This research has been conducted using the UK Biobank Resource under Application Number 8614.

VITAL: National Institutes of Health (K05 CA154337).

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005.

ASTERISK: We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students.

CCFR: The Colon CFR graciously thanks the generous contributions of their study participants, dedication of study staff, and the financial support from the U.S. National Cancer Institute, without which their important registry would not exist. The authors would like to thank the study participants and staff of the Seattle Colon Cancer Family Registry and the Hormones and Colon Cancer study (CORE Studies).

CLUE II: We thank the participants of Clue II and appreciate the continued efforts of the staff at the Johns Hopkins George W. Comstock Center for Public Health Research and Prevention in the conduct of the Clue II Cohort Study. Cancer data were provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data are also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

CORSA: We kindly thank all individuals who agreed to participate in the CORSA study. Furthermore, we thank all cooperating physicians and students and the Biobank Graz of the Medical University of Graz.

CPS-II: The authors express sincere appreciation to all Cancer Prevention Study-II participants, and to each member of the study and biospecimen management group. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries and cancer registries supported by the National Cancer Institute's Surveillance Epidemiology and End Results Program. The authors assume full responsibility for all analyses and interpretation of results. The views expressed here are those of the authors and do not necessarily represent the American Cancer Society or the American Cancer Society—Cancer Action Network.

Czech Republic CCS: We are thankful to all clinicians in major hospitals in the Czech Republic, without whom the study would not be practicable. We are also sincerely grateful to all patients participating in this study.

DACHS: We thank all participants and cooperating clinicians, and everyone who provided excellent technical assistance.

EDRN: We acknowledge all contributors to the development of the resource at University of Pittsburgh School of Medicine, Department of Gastroenterology, Department of Pathology, Hepatology and Nutrition and Biomedical Informatics.

EPIC: 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.

Harvard cohorts: The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We acknowledge Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital as home of the NHS. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. Central registries may also be supported by state agencies, universities, and cancer centers. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming. The authors assume full responsibility for analyses and interpretation of these data.

Kentucky: We would like to acknowledge the staff at the Kentucky Cancer Registry.

LCCS: We acknowledge the contributions of all who conducted this study which was originally reported as 10.1093/carcin/24.2.275.

NCCCS I & II: We would like to thank the study participants, and the NC Colorectal Cancer Study staff.

PLCO: The authors thank the PLCO Cancer Screening Trial screening centre investigators and the staff from Information Management Services Inc and Westat Inc. Most importantly, we thank the study participants for their contributions that made this study possible.

PPS3 and PPS4 Polyp Prevention Study trials: We would like to thank all of the study participants, investigators and staff.

Cancer incidence data have been provided by the District of Columbia Cancer Registry, Georgia Cancer Registry, Hawaii Cancer Registry, Minnesota Cancer Surveillance System, Missouri Cancer Registry, Nevada Central Cancer Registry, Pennsylvania Cancer Registry, Texas Cancer Registry, Virginia Cancer Registry, and Wisconsin Cancer Reporting System. All are supported in part by funds from the Center for Disease Control and Prevention, National Program for Central Registries, local states or by the National Cancer Institute, Surveillance, Epidemiology, and End Results program. The results reported here and the conclusions derived are the sole responsibility of the authors.

SEARCH: We thank the SEARCH team.

SELECT: We thank the research and clinical staff at the sites that participated on SELECT study, without whom the trial would not have been successful. We are also grateful to the 35,533 dedicated men who participated in SELECT.

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2025.105964.

References

1 Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249.
2 Morgan E, Arnold M, Gini A, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut*. 2023;72(2):338–344.
3 World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. Diet, nutrition, physical activity and colorectal cancer. Available at: dietandcancerreport.org.
4 O'Sullivan DE, Sutherland RL, Town S, et al. Risk factors for early-onset colorectal cancer: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2022;20(6):1229–1240.e5.
5 Yang T, Li X, Montazeri Z, et al. Gene-environment interactions and colorectal cancer risk: An umbrella review of systematic reviews and meta-analyses of observational studies. *Int J Cancer*. 2019;145(9):2315–2329.
6 Carreras-Torres R, Kim AE, Lin Y, et al. Genome-wide interaction study with smoking for colorectal cancer risk identifies novel genetic loci related to tumor suppression, inflammation, and immune response. *Cancer Epidemiol Biomarkers Prev*. 2023;32(3):315–328.
7 Jordahl KM, Shcherbina A, Kim AE, et al. Beyond GWAS of colorectal cancer: evidence of interaction with alcohol consumption and putative causal variant for the 10q24.2 region. *Cancer Epidemiol Biomarkers Prev*. 2022;31(5):1077–1089.
8 Tian Y, Kim AE, Bien SA, et al. Genome-wide interaction analysis of genetic variants with menopausal hormone therapy for colorectal cancer risk. *J Natl Cancer Inst*. 2022;114(8):1135–1148.
9 Dimou N, Kim AE, Flanagan O, et al. Probing the diabetes and colorectal cancer relationship using gene - environment interaction analyses. *Br J Cancer*. 2023;129:511–520.
10 Papadimitriou N, Kim A, Kawaguchi ES, et al. Genome-wide interaction study of dietary intake of fibre, fruits, and vegetables with risk of colorectal cancer. *EBioMedicine*. 2024;104:105146.
11 Stern MC, Sanchez Mendez J, Kim AE, et al. Genome-wide gene-environment interaction analyses to understand the relationship between red meat and processed meat intake and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2024;33(3):400–410.
12 Bouras E, Kim AE, Lin Y, et al. Genome-wide interaction analysis of folate for colorectal cancer risk. *Am J Clin Nutr*. 2023;118(5):881–891.
13 Aglago EK, Kim A, Lin Y, et al. A genetic locus within the FMN1/GREM1 gene region interacts with body mass index in colorectal cancer risk. *Cancer Res*. 2023;83(15):2572–2583.
14 Jeon J, Du M, Schoen REet al, Colorectal Transdisciplinary Study and Genetics and Epidemiology of Colorectal Cancer Consortium. Determining risk of colorectal cancer and starting age of screening based on lifestyle, environmental, and genetic factors. *Gastroenterology*. 2018;154(8):2152–2164.e19.
15 Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet*. 2019;51(1):76–87.
16 Peters U, Jiao S, Schumacher FRet al, Colon Cancer Family Registry and the Genetics and Epidemiology of Colorectal Cancer Consortium. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. *Gastroenterology*. 2013;144(4):799–807.e24.
17 1000 Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56–65.
18 Morrison J. GxEScanR: Run GWAS/GWEIS Scans Using Binary Dosage Files. R package version 2.0.2. Available at: https://CRAN.R-project.org/package=GxEScanR; 2020.
19 Cunningham F, Allen JE, Allen J, et al. Ensembl 2022. *Nucleic Acids Research*. 2021;50(D1):D988–D995.
20 Harrison PW, Amode MR, Austine-Orimoloye O, et al. Ensembl 2024. *Nucleic Acids Res*. 2024;52(D1):D891–D899.
21 Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–15550.
22 Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417–425.
23 Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739–1740.
24 Lin WY, Chen WJ, Liu CM, et al. Adaptive combination of Bayes factors as a powerful method for the joint analysis of rare and common variants. *Sci Rep*. 2017;7(1):13858.
25 Wellcome Trust Case Control Consortium. 000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–678.
26 Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–287.
27 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Ser B*. 1995;57(1):289–300.
28 Chen Y, Verbeek FJ, Wolstencroft K. Establishing a consensus for the hallmarks of cancer based on gene ontology and pathway annotations. *BMC Bioinformatics*. 2021;22(1):178.
29 Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646–674.
30 Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discov*. 2022;12(1):31–46.
31 Fernandez-Rozadilla C, Timofeeva M, Chen Z, et al. Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of European and East Asian ancestries. *Nat Genet*. 2023;55(1):89–99.
32 Ochoa D, Hercules A, Carmona M, et al. Open Targets Platform: supporting systematic drug–target identification and prioritisation. *Nucleic Acids Res*. 2021;49(D1):D1302–D1310.
33 R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available at: https://www.R-project.org/.
34 Huyghe JR, Harrison TA, Bien SA, et al. Genetic architectures of proximal and distal colorectal cancer are partly distinct. *Gut*. 2021;70(7):1325–1334.
35 Wang H, Burnett T, Kono S, et al. Trans-ethnic genome-wide association study of colorectal cancer identifies a new susceptibility locus in VTI1A. *Nat Commun*. 2014;5:4613.
36 Lu Y, Kweon SS, Tanikawa C, et al. Large-Scale Genome-Wide Association Study of East Asians identifies loci associated with risk for colorectal cancer. *Gastroenterology*. 2019;156(5):1455–1466.
37 Koveitypour Z, Panahi F, Vakilian M, et al. Signalling pathways involved in colorectal cancer progression. *Cell Biosci*. 2019;9:97.
38 Wang X, Amitay E, Harrison TA, et al. Association between smoking and molecular subtypes of colorectal cancer. *JNCI Cancer Spectr*. 2021;5(4):pkab056.
39 Dixon ZA, Nicholson L, Zeppetzauer M, et al. CREBBP knockdown enhances RAS/RAF/MEK/ERK signalling in RAS pathway mutated acute lymphoblastic leukemia but does not modulate chemotherapeutic response. *Haematologica*. 2017;102(4):736–745.
40 Ben Jemii N, Tounsi-Kettiti H, Yaiche H, et al. Dysregulated PDGFR alpha expression and novel somatic mutations in colorectal cancer: association to RAS wild type status and tumor size. *J Transl Med*. 2020;18(1):440.
41 Wang H, Li ZW, Ou Q, et al. NTRK fusion positive colorectal cancer is a unique subset of CRC with high TMB and microsatellite instability. *Cancer Med*. 2022;11(13):2541–2549.
42 Shah YM, van den Brink GR. c-Kit as a novel potential therapeutic target in colorectal cancer. *Gastroenterology*. 2015;149(3):534–537.

43 Phan NN, Wang CY, Chen CF, Sun Z, Lai MD, Lin YC. Voltage-gated calcium channels: novel targets for cancer therapy. *Oncol Lett*. 2017;14(2):2059–2074.
44 Donohoe F, Wilkinson M, Baxter E, Brennan DJ. Mitogen-activated protein kinase (MAPK) and obesity-related cancer. *Int J Mol Sci*. 2020;21(4):1241.
45 Dong Z, Lei X, Kujawa SA, Bolu N, Zhao H, Wang C. Identification of core gene in obese type 2 diabetes patients using bioinformatics analysis. *Adipocyte*. 2021;10(1):310–321.
46 Wong RSY. Role of Nonsteroidal Anti-Inflammatory Drugs (NSAIDs) in cancer prevention and cancer promotion. *Adv Pharmacol Sci*. 2019;2019:3418975.
47 Wang D, Dubois RN. The role of COX-2 in intestinal inflammation and colorectal cancer. *Oncogene*. 2010;29(6):781–788.
48 Dannenberg AJ, Altorki NK, Boyle JO, et al. Cyclo-oxygenase 2: a pharmacological target for the prevention of cancer. *Lancet Oncol*. 2001;2(9):544–551.
49 Gurpinar E, Grizzle WE, Piazza GA. COX-Independent mechanisms of cancer chemoprevention by anti-inflammatory drugs. *Front Oncol*. 2013;3:181.
50 Aguirre-Ducler A, Gianino N, Villarroel-Espindola F, et al. Tumor cell SYK expression modulates the tumor immune microenvironment composition in human cancer via TNF-α dependent signalling. *J Immunother Cancer*. 2022;10(7):e005113.
51 Geindreau M, Ghiringhelli F, Bruchard M. Vascular endothelial growth factor, a key modulator of the anti-tumor immune response. *Int J Mol Sci*. 2021;22(9):4871.
52 Open Targets Platform: colorectal neoplasm. Available at: https://platform.opentargets.org/. Accessed June 28, 2024.
53 World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. The cancer process. Available at: dietandcancerreport.org.
54 Murphy N, Newton CC, Song M, et al. Body mass index and molecular subtypes of colorectal cancer. *J Natl Cancer Inst*. 2022;115(2):165–173.
55 Hu C, Chen X, Yao C, et al. Body mass index-associated molecular characteristics involved in tumor immune and metabolic pathways. *Cancer Metab*. 2020;8(1):21.
56 Gonzalez DM, Medici D. Signalling mechanisms of the epithelial-mesenchymal transition. *Sci Signal*. 2014;7(344):re8.
57 Vu T, Jin L, Datta PK. Effect of cigarette smoking on epithelial to mesenchymal transition (EMT) in lung cancer. *J Clin Med*. 2016;5(4):44.
58 Pellegata NS, Berriel Diaz M, Rohm M, Herzig S. Obesity and cancer—Extracellular matrix, angiogenesis, and adrenergic signalling as unusual suspects linking the two diseases. *Cancer Metastasis Rev*. 2022;41(3):517–547.
59 Li J, Xu R. Obesity-associated ECM remodeling in cancer progression. *Cancers*. 2022;14(22):5684.
60 Hursting SD, Digiovanni J, Dannenberg AJ, et al. Obesity, energy balance, and cancer: new opportunities for prevention. *Cancer Prev Res*. 2012;5(11):1260–1272.
61 Scheurlen KM, Snook DL, Alfieri T, et al. Obesity hormones and itaconate mediating inflammation in human colon cancer cells - Another lead to early-onset colon cancer? *Heliyon*. 2023;9(2):e13132.
62 Cottam MA, Caslin HL, Winn NC, Hasty AH. Multiomics reveals persistence of obesity-associated immune cell phenotypes in adipose tissue during weight loss and weight regain in mice. *Nat Commun*. 2022;13(1):2950.
63 Shoelson SE, Lee J, Goldfine AB. Inflammation and insulin resistance. *J Clin Invest*. 2006;116(7):1793–1801.
64 Taylor EB. The complex role of adipokines in obesity, inflammation, and autoimmunity. *Clin Sci*. 2021;135(6):731–752.
65 Harding AT, Heaton NS. The impact of estrogens and their receptors on immunity and inflammation during infection. *Cancers*. 2022;14(4):909.
66 Akl MR, Nagpal P, Ayoub NM, et al. Molecular and clinical profiles of syndecan-1 in solid and hematological cancer for prognosis and precision medicine. *Oncotarget*. 2015;6(30):28693–28715.
67 Rothwell PM, Cook NR, Gaziano JM, et al. Effects of aspirin on risks of vascular events and cancer according to bodyweight and dose: analysis of individual patient data from randomised trials. *The Lancet*. 2018;392(10145):387–399.
68 Wang X, Chan AT, Slattery ML, et al. Influence of Smoking, body mass index, and other factors on the preventive effect of nonsteroidal anti-inflammatory drugs on colorectal cancer risk. *Cancer Res*. 2018;78(16):4790–4799.
69 Drew DA, Goh G, Mo A, et al. Colorectal polyp prevention by daily aspirin use is abrogated among active smokers. *Cancer Causes Control*. 2016;27(1):93–103.
70 Ishikawa H, Mutoh M, Suzuki S, et al. The preventive effects of low-dose enteric-coated aspirin tablets on the development of colorectal tumours in Asian patients: a randomised trial. *Gut*. 2014;63(11):1755–1759.
71 Costa AR, Duarte AC, Costa-Brito AR, Gonçalves I, Santos CRA. Bitter taste signalling in cancer. *Life Sci*. 2023;315:121363.
72 Chen JD, Evans RM. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature*. 1995;377(6548):454–457.
73 Matsui T, Heidaran M, Miki T, et al. Isolation of a novel receptor cDNA establishes the existence of two PDGF receptor genes. *Science*. 1989;243(4892):800–804.
74 Wu FY, Yin RX. Recent progress in epigenetics of obesity. *Diabetol Metab Syndr*. 2022;14(1):171.
75 Su D, Wang X, Campbell MR, et al. Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. *PloS one*. 2016;11(12):e0166486.
76 Al-Waeli H, Nicolau B, Stone L, et al. Chronotherapy of nonsteroidal anti-inflammatory drugs may enhance postoperative recovery. *Sci Rep*. 2020;10(1):468.
77 Hower IM, Harper SA, Buford TW. Circadian rhythms, exercise, and cardiovascular health. *J Circadian Rhythms*. 2018;16:7.
78 Barber LE, VoPham T, White LF, Roy HK, Palmer JR, Bertrand KA. Circadian disruption and colorectal cancer incidence in Black Women. *Cancer Epidemiol Biomarkers Prev*. 2023;32(7):927–935.
79 Angstadt AY, Hartman TJ, Lesko SM, et al. The effect of UGT1A and UGT2B polymorphisms on colorectal cancer risk: haplotype associations and gene–environment interactions. *Genes Chromosomes Cancer*. 2014;53(6):454–466.