scientific reports

Check for updates

OPEN

Genome-wide interaction study of physical activity and genetic susceptibility on colorectal cancer using UK biobank data

Sooyoung Cho[1]✉ & Aesun Shin[1,2,3,4]

Colorectal cancer (CRC) risk is influenced by a complex interplay between genetic predisposition and lifestyle factors, such as physical activity (PA). We aimed to conduct a genome-wide interaction study (GWIS) to explore single nucleotide polymorphisms (SNPs), and genes modulated by PA on CRC risk using data from the UK Biobank. Among 272,270 eligible participants, 2,979 CRC cases were matched with 11,435 controls using a incidence density matching approach to avoid potential biases that may arise when using excessively large unmatched control groups, and to preserve comparability in the timing and distribution of exposure. PA was defined as whether individuals met the international criteria. We used conditional logistic regression models to assess the significance for the SNP x PA interaction on CRC, and we also performed gene-level analysis by aggregating the results of SNP-level analysis. Several SNPs showed nominal interaction signals with $p < 5 \times 10^{-6}$, including loci mapped to *ABI3*, *ZBTB16*, and *GABRB3*, though none reached significance after FDR correction. Interaction and main effects were often in opposite directions. At the gene and pathway levels, RNASEL, NSD1, and efferocytosis showed nominal signals, although none reached statistical significance after correction. Although we could not find associations that met the significance threshold after adjusting for multiple testing, these preliminary findings help us to understand the interplay between genes and lifestyle in CRC.

Keywords Colorectal cancer, Single nucleotide polymorphism, Physical activity

Abbreviations

BMI	Body mass index
CRC	Colorectal cancer
GWAS	Genome-wide associaion study
GWIS	Genome-wide interaction study
PA	Physical activity
SNP	Single nucleotide polymorphism
QQ	Plot quantile-quantile plot

Colorectal cancer (CRC) is one of the leading causes of cancer-related morbidity and mortality worldwide, accounting for 9.6% of all new cancer cases and 9.3% of all cancer deaths in 2022[1]. Its development involves a complex interplay of genetic predisposition and modifiable lifestyle factors[2]. Numerous studies highlight the association between physical activity (PA) and the decreased risk for CRC, suggesting that regular PA lowers inflammation, improves insulin sensitivity, and modulates gut motility, all of which may contribute to reduced carcinogenesis[3–6]. However, the extent to which these benefits are influenced by individual genetic susceptibility remains not fully understood.

Large-scale genome-wide association studies (GWAS) have identified over 100 genetic loci associated with CRC susceptibility, implicating biological pathways such as Wnt signaling, immune regulation, and cell cycle control[7,8]. These findings have improved our understanding of CRC heritability, but most GWAS have focused solely on main genetic effects without considering interactions with behavioral or environmental exposures.

[1]Genomic Medicine Institute, Medical Research Center, Seoul National University, Seoul, Korea. [2]Department of Preventive Medicine, Seoul National University College of Medicine, Seoul, Korea. [3]Integrated Major in Innovative Medical Science, Seoul National University College of Medicine, Seoul, Korea. [4]Cancer Research Institute, Seoul National University, Seoul, Korea. ✉email: ssooy7@snu.ac.kr

While candidate gene studies have provided valuable insights into specific pathways, they are inherently limited in scope and fail to capture the broader genetic landscape influencing CRC risk. For example, polymorphisms in *IL6* and *TNF* (inflammation regulation), *FTO* and *PPARG* (energy metabolism), and *ABCA1* (lipid transport) have been associated with PA-modulated effects on biomarkers such as C-reactive protein, obesity-related traits, and lipid profiles[9–12]. In addition, genes such as *PITX1*, a tumor suppressor associated with IGF-I pathways, and oxidative stress-related genes such as *CAT*, *GSTP1* and *MPO* have been shown to interact with PA to influence cancer risk and antioxidant capacity[13,14]. These studies have been conducted to measure outcomes such as inflammatory markers, adiposity indices and oxidative stress levels, illustrating how genetic predisposition interacts with lifestyle factors in shaping disease risk. However, their reliance on prior biological assumptions and limited genomic coverage restricts their utility in discovering novel interactions[15].

Genome-wide interaction study (GWIS) can provide a robust and exploratory framework for uncovering novel gene-environment interactions. This approach is particularly useful for identifying genetic variants and pathways that have not previously been reported to be associated with CRC, thereby expanding our understanding of the interplay between PA and genetic factors in addition to the conventional genome-wide association studies (GWAS) approach[16]. In the context of CRC, only a few GWIS have been conducted to date, and these have focused on alcohol consumption[17], NSAIDS[18], and diet[19]. This highlights the need for systematic investigation into gene-PA interactions.

In this study, we aimed to investigate the interaction between genetic susceptibility and PA on CRC risk at a genome-wide level using data from the UK Biobank, employing a nested case-control design to minimize potential biases associated with excessively large unmatched control groups, and to preserve temporal comparability between cases and controls.

Methods

Study population

We used data from the UK Biobank (application #94695), a prospective cohort study of over 500,000 participants aged 40–69 years at baseline between 2006 and 2010.

After excluding participants who withdrew their consent, we applied the following exclusion criteria: missing information on the year or month of birth (UK Biobank field Data Field IDs: 34, 52), physical activity (Field ID: 22035) or smoking status (Field ID: 20116); missing information for all of the following: deprivation index (Field ID: 189), body mass index (BMI; Field ID: 21001) and alcohol consumption status (Field ID: 20117); non-European genetic ancestry (Field ID: 22006); a diagnosis of any cancer prior to the baseline assessment (Field ID: 53, 40005, 40006); and a genotyping call rate of less than 99% (Field ID: 22005). After applying these criteria, a total of 272,270 participants remained eligible for analysis (Fig. 1). Controls were selected through incidence density sampling from the eligible study population and were required to meet the same exclusion criteria as applied to cases, including no history of cancer prior to baseline.

We conducted a nested case-control study using incidence density matching to evaluate the interaction between physical activity (PA) and genetic susceptibility to CRC, ensuring that the controls represented the same risk set as the cases and preserving the temporal structure of exposure while maintaining comparability in a time-sensitive context. Incident CRC cases ($n = 2{,}974$) were identified through linkage with national cancer registries. For each case, up to four controls ($n = 11{,}424$) were selected from participants who were at risk at the time of case diagnosis. Controls were matched on sex, age at recruitment (± 5 years), smoking status and follow-up duration (± 6 months).

Genotyping data

We accessed genomic data provided by the UK Biobank, generated using the Affymetrix UK BiLEVE and UK Biobank Axiom Array platforms. These datasets contained over 800,000 single nucleotide polymorphisms (SNPs). Quality control procedures were applied to retain SNPs that met the following criteria: genotyping call rates > 99%, Hardy-Weinberg equilibrium p-values > 1×10^{-6}, and minor allele frequencies > 0.03. After filtering, 409,059 SNPs were included in the analysis.

We limited our analyses to directly genotyped SNPs, rather than imputed variants. Given the exploratory nature of the GWIS, we deliberately adopted a conservative approach. Although genotype imputation is widely used and generally reliable, interaction models are more susceptible to uncertainty in imputation, especially when modeling subtle gene-environment interactions. By focusing on high-confidence genotyped variants, we aimed to improve the robustness and interpretability of the results, even if it meant reducing genomic coverage.

Statistical analysis

We investigated the interaction between genetic variants and PA on CRC risk using conditional logistic regression models. The interaction between genetic variants and PA was assessed using the p for interaction, derived from the statistical significance of the interaction term (SNP × PA) in the regression model under an additive genetic model. Physical activity was categorized based on the 2017 WHO guidelines[20], with sufficient PA defined as at least 150 min of moderate-intensity or 75 min of vigorous-intensity activity per week.

To account for potential confounding, we adjusted the models for the following variables: BMI group, alcohol drinking status, socioeconomic deprivation index, the first 30 genetic principal components (to adjust for population stratification; Field ID: 22009), and genotyping batch (Field ID: 22000). The first 30 PCs were used as they collectively explained approximately 84.3% of the total genetic variance. Matching variables (age, sex, and smoking status) were not included in the adjustment as they were already accounted for by the matched study design.

To compare the baseline characteristics of cases and controls while accounting for the matched study design, we used conditional logistic regression with likelihood ratio tests (LRTs) for each covariate. This approach

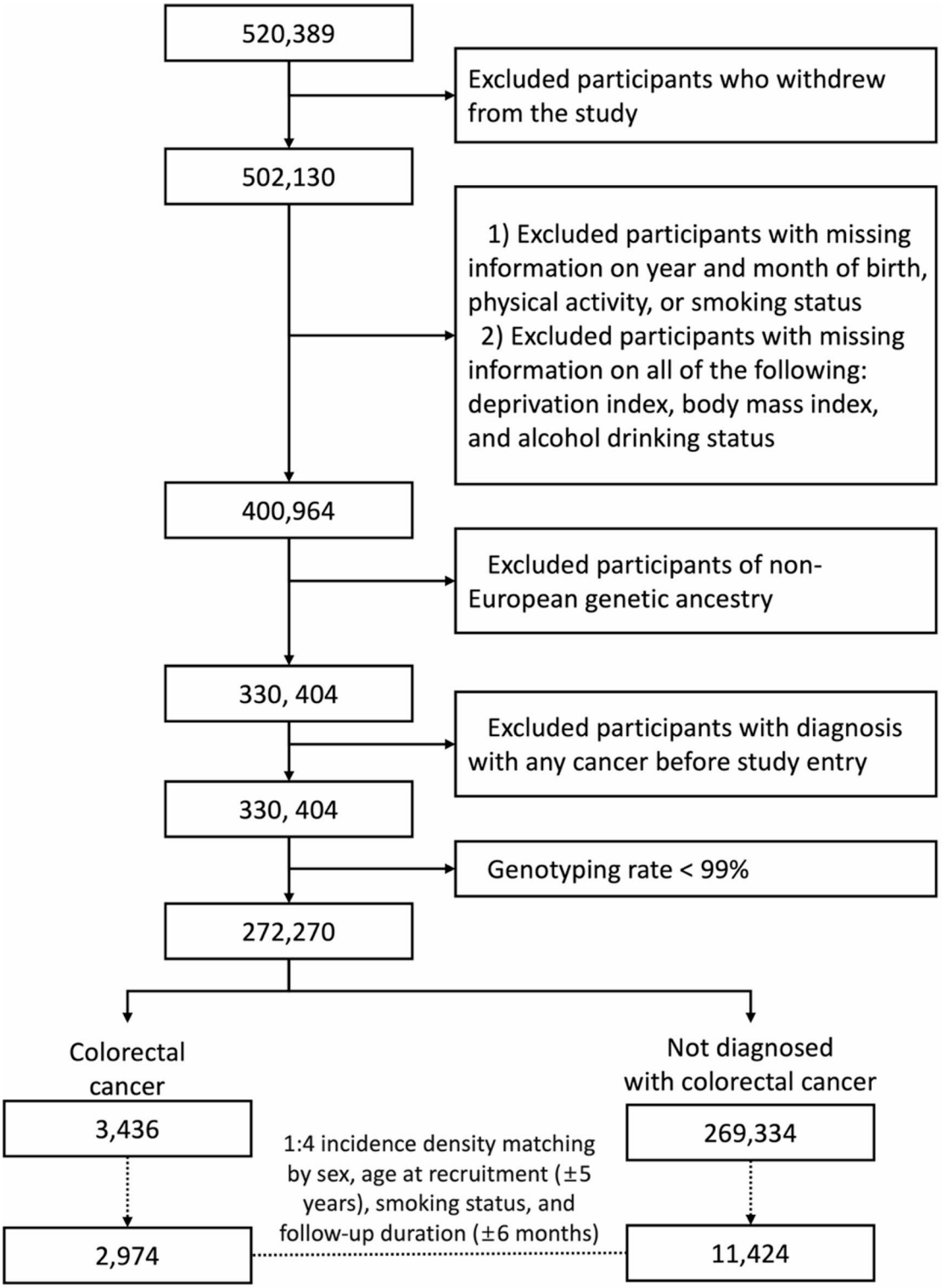

Fig. 1. Flowchart of study population selection based on incidence density matching using UK Biobank Data.

reflects the incidence density matching structure appropriately by conditioning on matched sets, and enables us to obtain a single overall P-value per variable. This is particularly useful for summarizing group differences in Table 1. Matching variables (age, sex and smoking status) were excluded from statistical testing as they were fixed by design and were not subject to comparison.

We performed pathway enrichment analysis using MAGMA (Multi-marker Analysis of GenoMic Annotation) version 1.10[21], following a three-step approach: SNP annotation to genes, gene-level association analysis, and pathway-level analysis. First, we annotated SNPs to genes based on their physical location using the NCBI37.3

Characteristics	Colorectal cancer cases	Controls	*P*-value
N	2,974	11,424	
Sex			
Male	1,793 (60.3)	6,943 (60.8)	
Female	1,181 (39.7)	4,481 (39.2)	
Age at study entry, median (IQR)	62.4 (9.2)	63.1 (8.3)	
Smoking status			
Never	1,317 (44.3)	5,019 (43.9)	
Former	1,345 (45.2)	5,182 (45.4)	
Current	312 (10.5)	1,223 (10.7)	
Alcohol drinking status			< 0.001
Never	67 (2.3)	426 (3.7)	
Former	106 (3.6)	657 (5.8)	
Current	2,799 (94.1)	10,336 (90.5)	
Body mass index, kg/m^2			< 0.001
< 18.5	8 (0.3)	76 (0.7)	
18.5–24.9	777 (26.1)	3,130 (27.4)	
25–29.9	1,359 (45.7)	4,744 (41.5)	
30–34.9	607 (20.4)	2,296 (20.1)	
≥ 35	215 (7.2)	1,058 (9.3)	
Deprivation index, median (IQR)	-2.4 (3.7)	-2.02 (4.4)	< 0.001
Physical activity			< 0.001
Sufficient (≥ 150 min moderate or 75 min vigorous per week)	1,582 (53.2)	5,839 (51.1)	
Insufficient	1,392 (46.8)	5,585 (48.9)	

Table 1. Baseline characteristics of colorectal cancer cases and matched controls in the nested case-control study from UK Biobank. P-values were derived from likelihood ratio tests in conditional logistic regression models that accounted for the matched study design. The matching variables (age, sex and smoking status) were excluded from the analysis as they were fixed by design.

gene reference file. SNPs located within a 10 kb window upstream or downstream of each gene were included in this process, resulting in a dataset that linked SNPs to their corresponding genes. We then calculated gene-level p-values by aggregating SNP-level p-values using a multiple regression framework implemented in MAGMA[22]. To account for linkage disequilibrium between SNPs, we used the European reference panel from the 1000 Genomes Project (phase 3). Finally, we performed pathway-level analysis by aggregating gene-level p-values into predefined pathways based on KEGG annotations[23]. Using MAGMA's competitive testing framework, we compared the observed associations within each pathway to the genome-wide background distribution.

False discovery rate (FDR) correction was applied at the SNP level to account for multiple testing, using the Benjamini–Hochberg method. No variants passed the significance threshold of an FDR-adjusted p-value of less than 0.05. Consequently, the top 10 variants with the lowest FDR-adjusted p-values were reported and interpreted as exploratory findings. FDR-adjusted p-values were also calculated for gene- and pathway-level analyses conducted using MAGMA. The full results, including the FDR-adjusted p-values, are presented in Tables S1 and S2.

Data preprocessing and quality control were performed using PLINK v2.0 and Python 3. The genome-wide interaction analysis was conducted using the clogit function in R version 4.3.1 to evaluate the interaction between SNPs and physical activity on colorectal cancer risk. Visualization was performed in R and included Manhattan plots, quantile–quantile (QQ) plots, and a volcano plot to display the direction and strength of interaction effects.

Results

Table 1 presents the baseline characteristics of colorectal cancer (CRC) cases ($n = 2{,}974$) and matched controls ($n = 11{,}424$) included in the final analysis. As the matching variables (age, sex, and smoking status) were fixed by design, statistical tests were not conducted for these variables. Statistically significant differences were observed in alcohol drinking status ($p < 0.001$), body mass index (BMI) category ($p < 0.001$), deprivation index ($p < 0.001$), and physical activity levels ($p < 0.001$). A slightly higher proportion of controls met the WHO guidelines for sufficient physical activity compared to cases (53.2% vs. 51.1%).

In the genome-wide interaction analysis, we assessed the interaction between genotyped SNPs and physical activity on CRC risk. No SNPs reached statistical significance after correction for multiple testing (FDR < 0.05). Table 2 summarizes the top 10 SNPs with the lowest FDR-adjusted p-values. The variant rs61856638 in the *ABI3* gene showed the strongest signal ($p = 1.11 \times 10^{-6}$; FDR-adjusted $p = 0.44$), followed by rs8043440 in *GABRB3* ($p = 2.16 \times 10^{-6}$; FDR-adjusted = 0.44) and rs1672718 in *ZBTB16* ($p = 4.62 \times 10^{-6}$; FDR-adjusted = 0.63). Several

SNP	Position	A > a	MAF	Main effect	Interaction effect		Mapped gene
				β (SE)	β (SE)	$P_{interaction}$	
rs1672718	11:113951186	C > G	0.42	0.14 (0.04)	− 0.29 (0.06)	2.16E-06	*ZBTB16*
rs61856638	10:88126072	A > G	0.15	0.26 (0.06)	− 0.40 (0.08)	1.11E-06	*ABI3*
rs8043440	15:26973086	C > T	0.12	0.27 (0.06)	− 0.42 (0.09)	4.62E-06	*GABRB3*
rs2300161	1:71436256	T > C	0.22	0.18 (0.05)	− 0.31 (0.07)	2.19E-05	–
rs2865162	20:39116791	A > G	0.08	0.12 (0.08)	− 0.46 (0.11)	2.86E-05	–
rs28684504	5:176727438	T > C	0.12	0.19 (0.06)	− 0.38 (0.09)	3.33E-05	–
rs34173893	1:94907022	A > AT	0.30	− 0.13 (0.05)	0.27 (0.06)	3.04E-05	–
rs3733875	5:176637240	T > G	0.11	0.20 (0.06)	− 0.38 (0.09)	3.45E-05	–
rs3893882	19:2409920	T > C	0.04	0.25 (0.10)	− 0.61 (0.14)	2.55E-05	–
rs2865162	20:39116791	A > G	0.08	0.12 (0.08)	− 0.46 (0.11)	2.86E-05	–

Table 2. Top 10 SNPs with the lowest FDR-adjusted p-values in the genome-wide interaction analysis of physical activity and colorectal cancer risk. None of the SNPs passed the FDR significance threshold (FDR < 0.05). Genomic coordinates are based on NCBI Build 37 (hg19). The A > a column represents the reference allele (major allele) and alternate allele (minor allele) for each SNP. The major and minor alleles are determined based on allele frequency within the study population included in the analyses, with the minor allele being the less frequent of the two. The interaction p-value ($P_{interaction}$) was derived from conditional logistic regression, assessing the interaction between SNP genotype and physical activity in colorectal cancer risk. The model was adjusted for deprivation index, body mass index (BMI), alcohol consumption, the first ten principal components of ancestry, and genotyping batch. The interaction significance was evaluated based on the p-value of the interaction term (SNP × physical activity) in the model.

of these SNPs showed moderate interaction effect sizes, though none surpassed the FDR-corrected significance threshold. The main effects of these SNPs on CRC risk were generally opposite the interaction terms. While none of these findings were statistically significant, this pattern may indicate potential interactions between genetic variation and physical activity that need to be investigated further.

Figure 2 shows the QQ plot of observed versus expected p-values, which closely followed the null distribution. Figure 3 displays the Manhattan plot of interaction p-values across the genome. No locus exceeded the genome-wide significance threshold, but several SNPs showed suggestive signals. Figure 4 illustrates the volcano plot highlighting the direction and magnitude of interaction effects, with the top 10 SNPs (based on FDR-adjusted p-values) marked.

A gene-level analysis was performed using MAGMA, which annotated 15,956 genes (Table S1). Although none of the genes passed the FDR-adjusted $p < 0.05$ threshold, several demonstrated low nominal p-values and may be of potential interest. These included *PTGFR* ($p = 7.48 \times 10^{-5}$; FDR-adjusted $p = 0.60$), *RNASEL* ($p = 7.76 \times 10^{-5}$; FDR-adjusted $p = 0.60$), *NSD1* ($p = 1.12 \times 10^{-4}$; FDR-adjusted $p = 0.60$) and *PTGER3* ($p = 1.83 \times 10^{-4}$; FDR-adjusted $p = 0.65$). While these results did not exceed FDR corrected threshold, they suggest candidate loci that may modulate CRC risk in relation to physical activity.

In the pathway-level analysis based on KEGG annotations, no pathways reached statistical significance after FDR correction. However, several pathways ranked among the top results based on their unadjusted p-values. These included platinum drug resistance ($p = 0.0083$; FDR-adjusted $p = 0.85$), heparan sulfate/heparin biosynthesis ($p = 0.0084$; FDR-adjusted $p = 0.85$), efferocytosis ($p = 0.0148$; FDR-adjusted $p = 0.85$), and transcriptional misregulation in cancer ($p = 0.0168$; FDR-adjusted $p = 0.85$). Inflammation-related pathways such as NF-κB signaling and Notch signaling also appeared among the top-ranked findings. While these pathways did not meet the significance threshold after multiple testing correction, they may offer biologically plausible leads for future investigation. Full pathway-level results are presented in Table S2.

Discussion

In this genome-wide interaction study, we investigated whether physical activity modifies genetic susceptibility to CRC. After applying false discovery rate correction, no variants, genes, or pathways reached statistical significance. These results highlight the difficulty of identifying gene-environment interactions in complex diseases and emphasize the exploratory nature of our analysis. Although genome-wide significant interaction signals cannot be observed, this may be partly due to the lack of a significant association between physical activity and CRC risk in this cohort (odds ratio [95% confidence interval] 1.07 [0.99–1.16] in the multivariable model). Limited power to detect modest interaction effects and potential exposure misclassification may also have contributed.

Some of the variants with relatively low interaction p-values were located within genes that may have biological relevance to CRC, particularly in the context of immune regulation or epigenetic control. *ABI3* and *ZBTB16* are both involved in immune cell signaling and differentiation, with *ZBTB16* previously linked to colorectal tumorigenesis through modulation of Wnt signaling and inflammatory pathways[24]. *RNASEL*, a gene involved in interferon-mediated antiviral responses, has been linked to cancer susceptibility in previous studies[25]. Although a recent study reported no significant association between *RNASEL* variants and colorectal adenoma risk[26], its relevance to colorectal cancer may vary across stages of tumor development. *NSD1* encodes

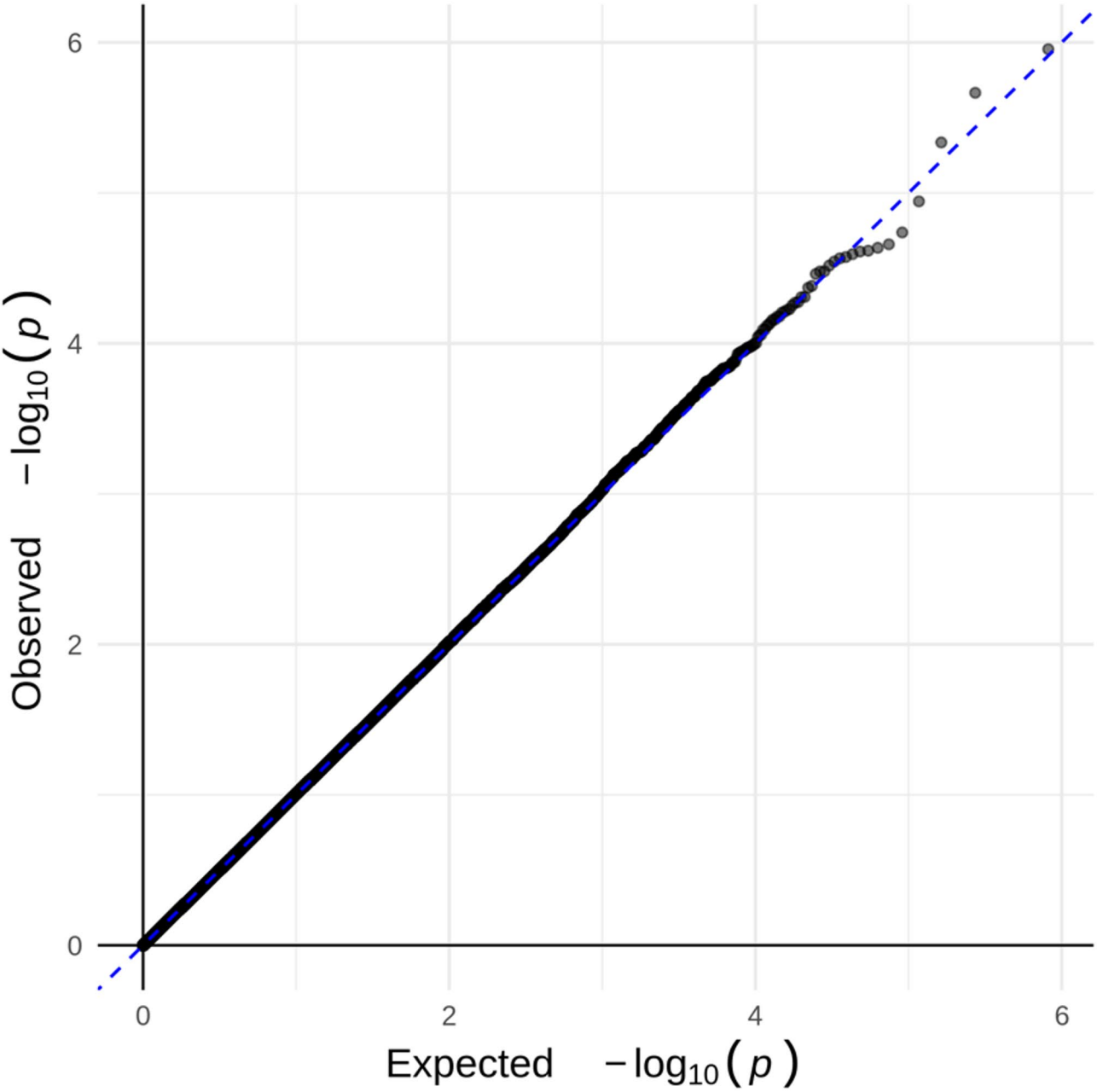

Fig. 2. Quantile–quantile plot of genome-wide SNP × physical activity interaction p-values.

a histone methyltransferase involved in epigenetic regulation and chromatin remodeling. Although its role in colorectal cancer is not well established, *NSD1* has been implicated in other malignancies and overgrowth syndromes through altered transcriptional regulation[27,28]. *GABRB3* has been associated with neurological and psychiatric conditions, including epilepsy[29], autism spectrum disorders[30], and bipolar disorder[31]. Although its role in colorectal cancer remains unclear, GABAergic signaling has been investigated in relation to epithelial cell function[32] and intestinal homeostasis[33]. While none of these findings reached statistical significance after multiple testing correction, they may offer tentative biological clues that merit further investigation.

In the pathway-level analysis, no KEGG-defined pathways reached statistical significance after FDR correction. Nonetheless, several pathways had relatively low unadjusted p-values and were ranked among the top results, which may offer preliminary insights for future hypothesis-driven research. Among them, the platinum drug resistance pathway includes genes involved in DNA repair mechanisms and apoptosis regulation—cellular processes that have been associated with colorectal cancer progression and may be modulated by physical activity–induced changes in oxidative stress and cellular stress response pathways[34,35]. The heparan sulfate/heparin biosynthesis pathway is related to glycosaminoglycan metabolism, which can affect cell adhesion and extracellular matrix interactions. These factors influence epithelial integrity and tumor invasion, and may also respond to biomechanical or hormonal changes induced by regular physical activity[36].

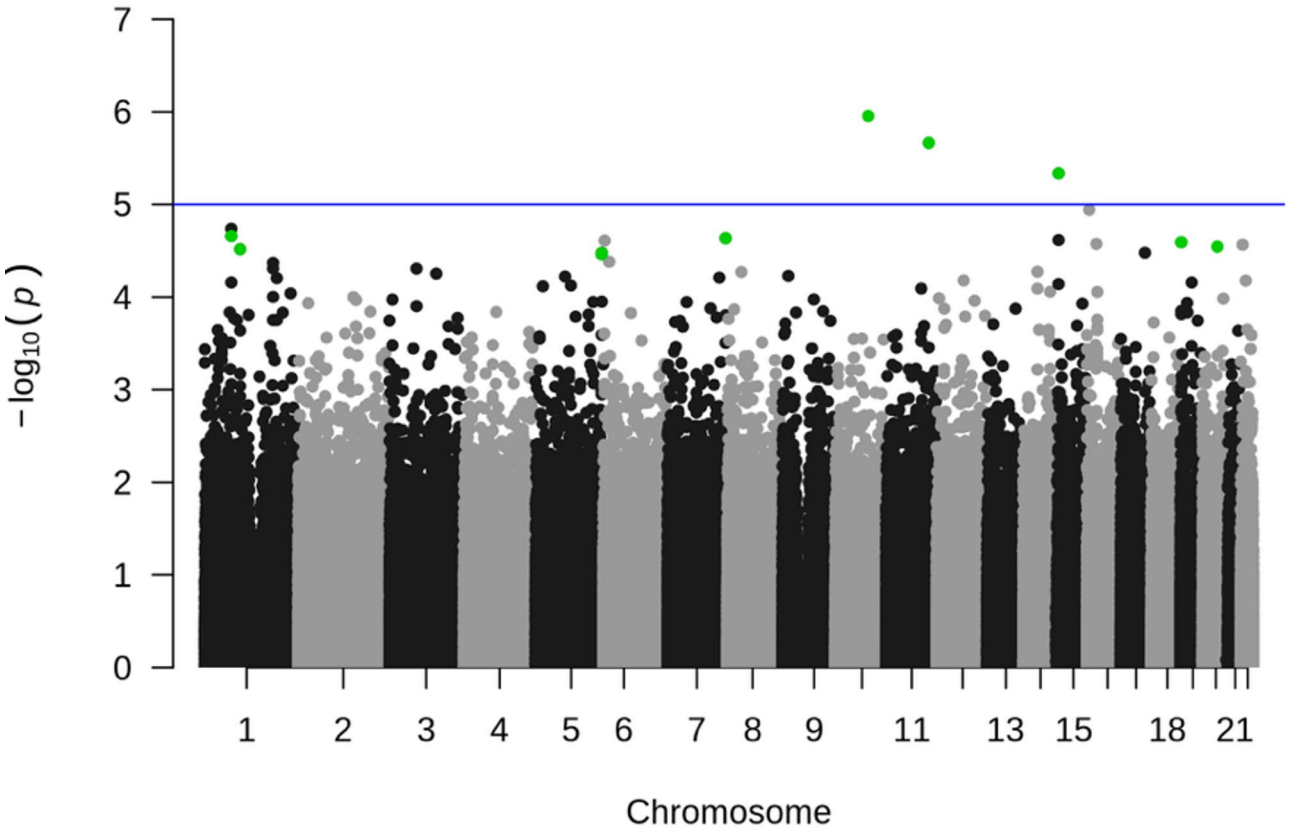

Fig. 3. Manhattan plot of genome-wide SNP × physical activity interaction on colorectal cancer risk. SNPs with the top 10 smallest FDR-adjusted p-values are highlighted in green.

Efferocytosis, the process by which phagocytes clear apoptotic cells, is essential for resolving inflammation and maintaining immune tolerance. Disruption of this process has been linked to chronic inflammation and the formation of tumor-promoting microenvironments in the colon[37,38]. Physical activity is known to influence systemic inflammatory tone, which may interact with cell death and efferocytic pathways in shaping CRC risk[39]. Lastly, transcriptional misregulation in cancer encompasses a diverse set of genes frequently altered in tumorigenesis, including those related to cell cycle control and differentiation. Though broad, this category may capture regulatory pathways influenced by both genetic variation and lifestyle exposures[40]. Although these pathways did not reach FDR-adjusted significance and should be interpreted with caution, the convergence of inflammation-, repair-, and differentiation-related processes among the top results may suggest biological pathways through which physical activity and genetic variation could jointly influence CRC development.

This study has several limitations. The sample size may have limited power to detect modest gene–environment interactions, and physical activity was self-reported, raising the possibility of measurement error. Although major covariates were adjusted for, residual confounding cannot be excluded. Despite these constraints, the use of incidence density matching helped minimize time-related bias and ensured appropriate comparability between cases and controls. Analyses were restricted to directly genotyped variants to reduce uncertainty from imputation, and population structure was controlled using principal components. These design features strengthen the internal validity of the findings, even in the absence of statistically significant associations.

This exploratory genome-wide interaction analysis did not identify statistically significant associations between physical activity and genetic variants in colorectal cancer. Nonetheless, several variants and pathways showed nominal signals that may inform future hypotheses. Although these findings require cautious interpretation, the use of a matched design, direct genotyping, and multi-level analysis supports the integrity of the results and their potential value as a starting point for further research.

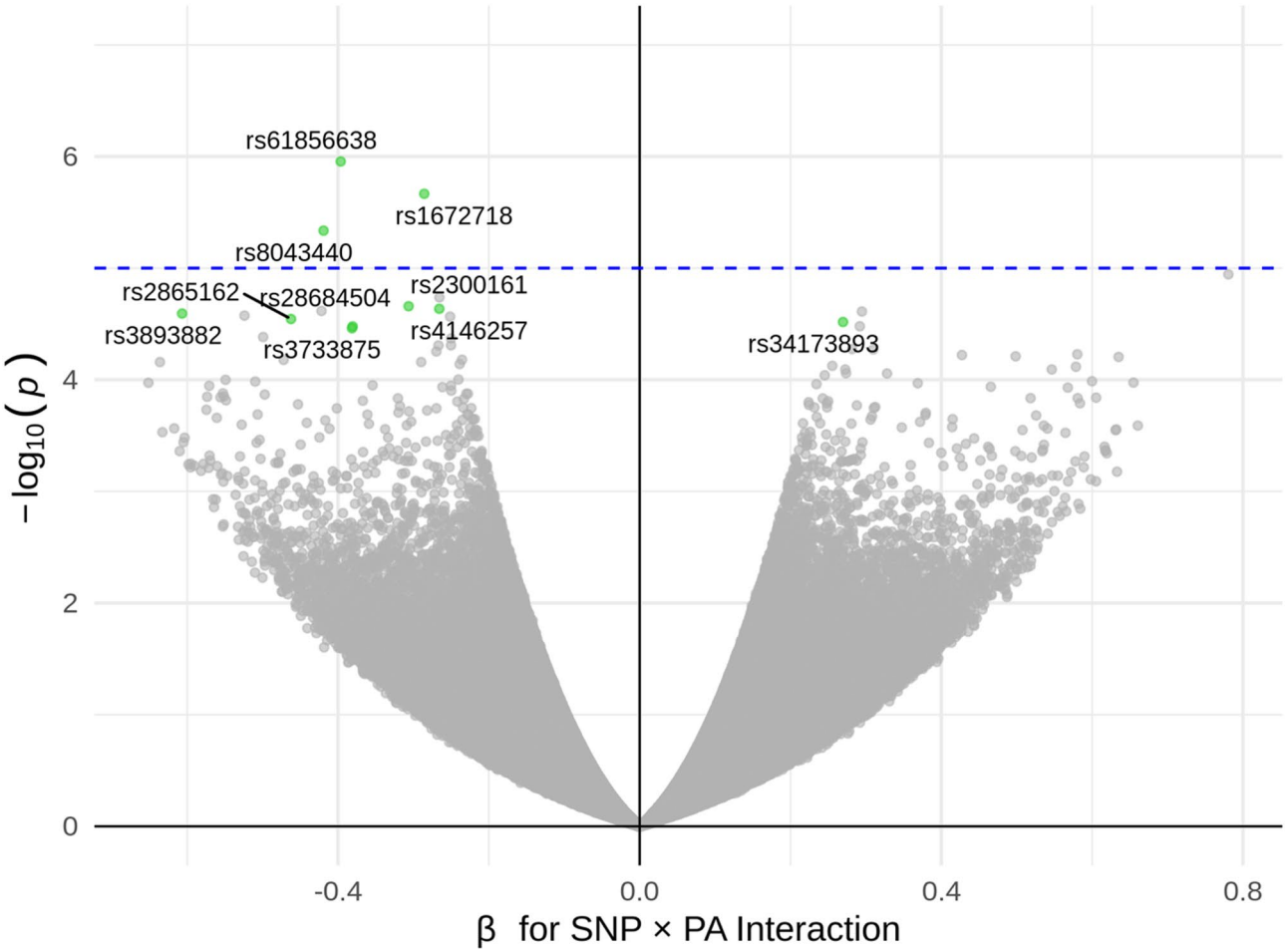

Fig. 4. Volcano plot of SNP × physical activity interaction effects. SNPs with the top 10 smallest FDR-adjusted p-values are highlighted in green.

Data availability

The study used data from the UK Biobank (https://www.ukbiobank.ac.uk). Researchers can access UK Biobank data if they apply and follow the rules. The dataset analyzed in this study can be accessed via the UK Biobank (application number: 94695). SNP-level summary statistics are not publicly posted but can be available from the corresponding author upon reasonable request.

Received: 13 March 2025; Accepted: 25 July 2025
Published online: 18 August 2025

References

1. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74**, 229–263. https://doi.org/10.3322/caac.21834 (2024).
2. Simonds, N. I. et al. Review of the Gene-Environment interaction literature in cancer: what do we know?? *Genet. Epidemiol.* **40**, 356–365. https://doi.org/10.1002/gepi.21967 (2016).
3. Thune, I. & Lund, E. Physical activity and risk of colorectal cancer in men and women. *Br. J. Cancer*. **73**, 1134–1140. https://doi.org/10.1038/bjc.1996.218 (1996).
4. Samad, A. K., Taylor, R. S., Marshall, T. & Chapman, M. A. A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. *Colorectal Dis.* **7**, 204–213. https://doi.org/10.1111/j.1463-1318.2005.00747.x (2005).
5. Howard, R. A. et al. Physical activity, sedentary behavior, and the risk of colon and rectal cancer in the NIH-AARP diet and health study. *Cancer Causes Control.* **19**, 939–953. https://doi.org/10.1007/s10552-008-9159-0 (2008).
6. Papadimitriou, N. et al. Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis. *Nat. Commun.* **11**, 597. https://doi.org/10.1038/s41467-020-14389-8 (2020).
7. Huyghe, J. R. et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat. Genet.* **51**, 76–87. https://doi.org/10.1038/s41588-018-0286-6 (2019).
8. Law, P. J. et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat. Commun.* **10**, 2154. https://doi.org/10.1038/s41467-019-09775-w (2019).
9. Andreasen, C. H. et al. Low physical activity accentuates the effect of the FTO rs9939609 polymorphism on body fat accumulation. *Diabetes* **57**, 95–101. https://doi.org/10.2337/db07-0910 (2008).

10. Kilpelainen, T. O. et al. The rs1800629 polymorphism in the TNF gene interacts with physical activity on the changes in C-reactive protein levels in the Finnish diabetes prevention study. *Exp. Clin. Endocrinol. Diabetes*. **118**, 757–759. https://doi.org/10.1055/s-0030-1249686 (2010).
11. Gunathilake, M. N. et al. Interaction between physical activity, PITX1 rs647161 genetic polymorphism and colorectal cancer risk in a Korean population: a case-control study. *Oncotarget* **9**, 7590–7603. https://doi.org/10.18632/oncotarget.24136 (2018).
12. Nishida, Y. et al. The interaction between ABCA1 polymorphism and physical activity on the HDL-cholesterol levels in a Japanese population[S]. *J. Lipid Res.* **61**, 86–94. https://doi.org/10.1194/jlr.P091546 (2020).
13. McCullough, L. E. et al. Polymorphisms in oxidative stress genes, physical activity, and breast cancer risk. *Cancer Causes Control.* **23**, 1949–1958. https://doi.org/10.1007/s10552-012-0072-1 (2012).
14. Salazar-Tortosa, D. F. et al. Association between polymorphisms and obesity-related phenotypes in European adolescents: influence of physical activity. *Pediatr. Res.* **93**, 2036–2044. https://doi.org/10.1038/s41390-022-02377-1 (2023).
15. Thomas, D. Gene–environment-wide association studies: emerging approaches. *Nat. Rev. Genet.* **11**, 259–272. https://doi.org/10.1038/nrg2764 (2010).
16. Li, P., Guo, M. Z., Wang, C. Y., Liu, X. Y. & Zou, Q. An overview of SNP interactions in genome-wide association studies. *Brief. Funct. Genomics*. **14**, 143–155. https://doi.org/10.1093/bfgp/elu036 (2015).
17. Jordahl, K. M. 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.* **31**, 1077–1089. https://doi.org/10.1158/1055-9965.Epi-21-1003 (2022).
18. Kim, A. E. et al. Abstract 820: NSAIDs and colorectal cancer risk: results from genome-wide interaction scans. *Cancer Res.* **81**, 820–820. https://doi.org/10.1158/1538-7445.Am2021-820 (2021).
19. Hoang, T., Cho, S., Choi, J. Y., Kang, D. & Shin, A. Genome-Wide interaction study of dietary intake and colorectal cancer risk in the UK biobank. *JAMA Netw. Open.* **7**, e240465–e240465. https://doi.org/10.1001/jamanetworkopen.2024.0465 (2024).
20. Bull, F. C. et al. World health organization 2020 guidelines on physical activity and sedentary behaviour. *Br. J. Sports Med.* **54**, 1451–1462. https://doi.org/10.1136/bjsports-2020-102955 (2020).
21. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219. https://doi.org/10.1371/journal.pcbi.1004219 (2015).
22. de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat. Rev. Genet.* **17**, 353–364. https://doi.org/10.1038/nrg.2016.29 (2016).
23. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677. https://doi.org/10.1093/nar/gkae909 (2024).
24. Tong, Y., Song, Y., Xia, C. & Deng, S. Theoretical and in silico analyses reveal MYC as a dynamic network biomarker in colon and rectal cancer. *Frontiers in Genetics* **11**. https://doi.org/10.3389/fgene.2020.555540 (2020).
25. Zhou, A., Molinaro, R. J., Malathi, K. & Silverman, R. H. Mapping of the human RNASEL promoter and expression in cancer and normal cells. *J. Interferon Cytokine Res.* **25**, 595–603. https://doi.org/10.1089/jir.2005.25.595 (2005).
26. Huang, B. Z. et al. Polymorphisms in genes related to inflammation and obesity and colorectal adenoma risk. *Mol. Carcinog.* **57**, 1278–1288. https://doi.org/10.1002/mc.22842 (2018).
27. Niikawa, N. Molecular basis of sotos syndrome. *Horm. Res.* **62** (Suppl 3), 60–65. https://doi.org/10.1159/000080501 (2004).
28. Romero, V. I., Arias-Almeida, B. & Aguiar, S. A. NSD1 gene evolves under episodic selection within primates and mutations of specific exons in humans cause sotos syndrome. *BMC Genom.* **23**, 849. https://doi.org/10.1186/s12864-022-09071-w (2022).
29. Hempelmann, A. et al. Lack of evidence of an allelic association of a functional GABRB3 exon 1a promoter polymorphism with idiopathic generalized epilepsy. *Epilepsy Res.* **74**, 28–32. https://doi.org/10.1016/j.eplepsyres.2006.12.001 (2007).
30. Kim, S. A., Kim, J. H., Park, M., Cho, I. H. & Yoo, H. J. Association of GABRB3 polymorphisms with autism spectrum disorders in Korean trios. *Neuropsychobiology* **54**, 160–165. https://doi.org/10.1159/000098651 (2006).
31. Papadimitriou, G. N. et al. GABA-A receptor beta3 and alpha5 subunit gene cluster on chromosome 15q11-q13 and bipolar disorder: a genetic association study. *Am. J. Med. Genet.* **105**, 317–320. https://doi.org/10.1002/ajmg.1354 (2001).
32. Li, Y., Xiang, Y. Y., Lu, W. Y., Liu, C. & Li, J. A novel role of intestine epithelial GABAergic signaling in regulating intestinal fluid secretion. *Am. J. Physiology-Gastrointestinal Liver Physiol.* **303**, G453–G460. https://doi.org/10.1152/ajpgi.00497.2011 (2012).
33. Auteri, M., Zizzo, M. G. & Serio, R. GABA and GABA receptors in the Gastrointestinal tract: from motility to inflammation. *Pharmacol. Res.* **93**, 11–21. https://doi.org/10.1016/j.phrs.2014.12.001 (2015).
34. Lord, C. J. & Ashworth, A. The DNA damage response and cancer therapy. *Nature* **481**, 287–294. https://doi.org/10.1038/nature10760 (2012).
35. Tryfidou, D. V., McClean, C., Nikolaidis, M. G. & Davison, G. W. DNA damage following acute aerobic exercise: A systematic review and Meta-analysis. *Sports Med.* **50**, 103–127. https://doi.org/10.1007/s40279-019-01181-y (2020).
36. Blackhall, F. H., Merry, C. L. R., Davies, E. J. & Jayson, G. C. Heparan sulfate proteoglycans and cancer. *Br. J. Cancer*. **85**, 1094–1098. https://doi.org/10.1054/bjoc.2001.2054 (2001).
37. Arandjelovic, S. & Ravichandran, K. S. Phagocytosis of apoptotic cells in homeostasis. *Nat. Immunol.* **16**, 907–917. https://doi.org/10.1038/ni.3253 (2015).
38. Terzić, J., Grivennikov, S., Karin, E. & Karin, M. Inflammation and colon cancer. *Gastroenterology* **138**, 2101–2114e2105. https://doi.org/10.1053/j.gastro.2010.01.058 (2010).
39. Pedersen, B. K. & Febbraio, M. A. Muscles, exercise and obesity: skeletal muscle as a secretory organ. *Nat. Reviews Endocrinol.* **8**, 457–465. https://doi.org/10.1038/nrendo.2012.49 (2012).
40. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674. https://doi.org/10.1016/j.cell.2011.02.013 (2011).

Author contributions

SC analyzed the data, generated the tables, and wrote the manuscript. AS contributed to data acquisition and commented on the study design. All authors reviewed and edited the manuscript before submission.

Funding information

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government. (MSIT) (No. NRF-2021R1A6A3A01088767 and 2022R1A2C1004608).

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

We utilized data obtained from the UK Biobank, a large-scale biomedical resource established under ethical guidelines. Ethics approval for the UK Biobank was granted by the North West Multi-centre Research Ethics Committee. All participants provided informed consent at the time of data collection by the UK Biobank. We accessed these data, including genotyping information, under approved data access permissions. The study was conducted in accordance with the principles outlined in the Declaration of Helsinki.

Consent for publication

This manuscript does not include any identifiable individual person's data. All data used in this study were anonymized and obtained under appropriate ethical and legal guidelines.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-13709-6.

Correspondence and requests for materials should be addressed to S.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.