

Exposure to air pollution increases susceptibility to ulcerative colitis through epigenetic alterations in CXCR2 and MHC class III region



Jie Chen,^{a,b,m} Han Zhang,^{a,m} Tian Fu,^{a,b,c,m} Jianhui Zhao,^a Jan Krzysztof Nowak,^d Rahul Kalla,^e Judith Wellens,^{f,g} Shuai Yuan,^h Alexandra Noble,^g Nicholas T. Ventham,^e Malcolm G. Dunlop,ⁱ Jonas Halfvarson,^j Ren Mao,^k Evropi Theodoratou,^{i,l,n,**} Jack Satsangi,^{g,n,***} and Xue Li^{a,n,*}



^aThe Second Affiliated Hospital and School of Public Health, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

^bDepartment of Gastroenterology, The Third Xiangya Hospital, Central South University, Changsha, China

^cDepartment of Gastroenterology, Affiliated Hangzhou First People's Hospital, Westlake University Medical College, Hangzhou, China

^dDepartment of Pediatric Gastroenterology and Metabolic Diseases, Poznan University of Medical Sciences, 60572, Poznan, Poland

^eMedical Research Council Centre for Inflammation Research, Queens Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom

^fKU Leuven Department of Chronic Diseases and Metabolism, Translational Research Center for Gastrointestinal Disorders (TARGID), Leuven, Belgium

^gTranslational Gastro-Intestinal Unit, Nuffield Department of Medicine, John Radcliffe Hospital, Oxford, UK

^hUnit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

ⁱCancer Research UK Edinburgh Centre, Medical Research Council Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

^jDepartment of Gastroenterology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

^kDepartment of Gastroenterology, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

^lCentre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

Summary

Background This study aims to confirm the associations of air pollution with ulcerative colitis (UC) and Crohn's disease (CD); to explore interactions with genetics and lifestyle; and to characterize potential epigenetic mechanisms.

Methods We identified over 450,000 individuals from the UK Biobank and investigated the relationship between air pollution and incident inflammatory bowel disease (IBD). Cox regression was utilized to calculate hazard ratios (HRs), while also exploring potential interactions with genetics and lifestyle factors. Additionally, we conducted epigenetic Mendelian randomization (MR) analyses to examine the association between air pollution-related DNA methylation and UC. Finally, our findings were validated through genome-wide DNA methylation analysis of UC, as well as co-localization and gene expression analyses.

Findings Higher exposures to NO_x (HR = 1.20, 95% CI 1.05–1.38), NO₂ (HR = 1.19, 95% CI = 1.03–1.36), PM_{2.5} (HR = 1.19, 95% CI = 1.05–1.36) and combined air pollution score (HR = 1.26, 95% CI = 1.11–1.45) were associated with incident UC but not CD. Interactions with genetic risk score and lifestyle were observed. In MR analysis, we found five and 22 methylated CpG sites related to PM_{2.5} and NO₂ exposure to be significantly associated with UC. DNA methylation alterations at CXCR2 and sites within the MHC class III region, were validated in genome-wide DNA methylation analysis, co-localization analysis and analysis of colonic tissue.

eBioMedicine

2024;110: 105443

Published Online xxx

<https://doi.org/10.1016/j.ebiom.2024.105443>

1016/j.ebiom.2024.105443

105443

Abbreviations: UC, Ulcerative colitis; DNA, Deoxyribonucleic acid; NO₂, Nitrogen dioxide; PM, Particulate matter; NO_x, Nitrogen oxides; HR, Hazard ratios; CI, Confidence interval; MR, Mendelian randomization; IBD, Inflammatory bowel disease; CD, Crohn's disease; EWAS, Epigenome-wide association studies; ICD, International Statistical Classification of Disease and Related Health Problems; BMI, Body mass index; IQR, Interquartile range; GoDMC, Genetics of DNA Methylation Consortium; mQTL, methylation quantitative trait loci; GWAS, Genome-wide association study; IVW, Inverse-variance weighted; CXCR2, CXC motif chemokine receptor 2; THIN, The health improvement network; EPIC, European Prospective Investigation into Cancer and Nutrition cohort; SNP, Single nucleotide polymorphism; DMP, Differentially methylated position; Chr, Chromosome; LUR, Land use regression; SD, Standard deviation; TDI, Townsend deprivation index; FDR, False discovery rate; PRS, Polygenic risk score; MHC, Major histocompatibility complex; HLA, Human leukocyte antigen; RCT, Randomized controlled trial

*Corresponding author.

**Corresponding author. Cancer Research UK Edinburgh Centre, Medical Research Council Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK.

***Corresponding author.

E-mail addresses: xueli157@zju.edu.cn (X. Li), e.theodoratou@ed.ac.uk (E. Theodoratou), jack.satsangi@ndm.ox.ac.uk (J. Satsangi).

^{††}Contributed equally as first authors.

^{†††}Contributed equally as senior last authors.

Interpretation We report a potential causal association between air pollution and UC, modified by lifestyle and genetic influences. Biological pathways implicated include epigenetic alterations in key genetic loci, including *CXCR2* and susceptible loci within MHC class III region.

Funding Xue Li was supported by the Natural Science Fund for Distinguished Young Scholars of Zhejiang Province (LR22H260001) and the National Nature Science Foundation of China (No. 82204019). ET was supported by the CRUK Career Development Fellowship (C31250/A22804) and the Research Foundation Flanders (FWO). JW was supported by Belgium by a PhD Fellowship strategic basic research (SB) grant (1S06023N). JKN was supported by the National Science Center, Poland (No. 2020/39/D/NZ5/02720). The IBD Character was supported by the European Union's Seventh Framework Programme [FP7] grant IBD Character (No. 2858546).

Copyright © 2024 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: Inflammatory bowel disease; Air pollution; DNA methylation; Mendelian randomization; MHC III region

Research in context

Evidence before this study

Environmental factors play a role in the pathogenesis of inflammatory bowel disease (IBD), with air pollution emerging as a potential environmental risk factor. A recent analysis has revealed an elevated risk of ulcerative colitis (UC) associated with exposure to various air pollutants.

Added value of this study

We confirm that exposure to NO_x , NO_2 , and $\text{PM}_{2.5}$ are associated with increased risk of UC. Five and 22 differentially methylated sites related to $\text{PM}_{2.5}$ and NO_2 were significantly associated with the risk of UC. Effects of DNA methylation alterations at *CXCR2* and loci within major histocompatibility complex (MHC) class III region were further validated.

Implications of all the available evidence

Our study presents an exciting research framework for future investigations aiming to elucidate the potential mechanisms of environment-IBD by integrating observational, genetic, and methylation evidence. Additionally, it underscores the significance of epigenetic modifications in the complex pathogenesis of UC. The discovery of epigenetic dysregulation linking air pollution to UC risk demonstrates the translational potential of these methodologies in identifying biomarkers and characterizing pathways for pharmacological interventions.

Introduction

Incidence rates of inflammatory bowel diseases (IBD), principally Crohn's disease (CD) and ulcerative colitis (UC), are increasing across all age groups.^{1,2} It is of particular note that the increase has been substantial in recent years¹ in densely populated urbanized areas of newly industrialized countries. Globally, the prevalence of IBD continues to rise, and is projected to soon reach 1% in certain countries, such as the UK.³ Whilst genetic factors are important in IBD pathogenesis,⁴ environmental factors must underlie the recent rise in IBD incidence.⁵

We have explored a number of potential environmental risk factors in recent analyses of the UK Biobank, a cohort of approximately 500,000 adults followed prospectively since recruitment from 2006 to 2010. In particular, we have reported in detail on associations of body mass index, physical activities, sleep,⁶ diet,^{7,8} and smoking⁹ with the risk of incident IBD, and reviewed gene-environmental interactions on IBD.¹⁰ Air pollution merits examination as a potential factor, given its increasing impact on the environment and public health over the past decades.¹¹ Epidemiological evidence has

emerged to suggest a potential association between air pollution and onset of IBD.^{12–16} Most relevant in the present context is the recent analysis conducted within the UK Biobank and reported by Li et al. providing significant evidence of positive associations between $\text{PM}_{2.5}$, PM_{10} , NO_2 , and NO_x exposure and the risk of UC.¹⁶ However, current knowledge is incomplete, notably in the evaluation of the interactions between air pollution, genetic susceptibility and lifestyle factors in modifying the risk of IBD. Numerous studies have indicated interactions of these factors in common diseases, such as lung cancer,¹⁷ mortality,¹⁸ mental disorders,^{19–21} rheumatoid arthritis,²² and abdominal aortic aneurysm.²³

Potential mechanisms involved in the effect of pollutants on disease susceptibility require exploration, and the epigenetic alterations are directly relevant. Deoxyribonucleic acid (DNA) methylation is the most widely investigated epigenetic change that could explain the impact of gene-environment interaction in disease development.²⁴ Altered methylation has been suggested as a potential mediator between air pollution and a series of chronic diseases, including metabolic syndrome,²⁵ cancer,²⁶ and adverse respiratory health outcomes.²⁶

While recent data implicate DNA methylation in the pathogenesis of IBD and the effect of smoking on IBD,^{9,24,27–31} its potential role in the association between air pollution and IBD has not yet been explored.

Evidence for the mechanistic importance of DNA methylation in this context has increased. A randomized controlled trial (RCT) study has found that exposure to PM_{2.5} is linked to increased levels of various inflammatory cytokines and DNA methylation in corresponding genes, and indoor air filtration can mitigate these negative effects.³² Mendelian randomization (MR) analysis is now accepted as a valuable complementary analytic approach for assessing the causality of associations,³³ using genetic variants allocated randomly during human gamete formation as instrument variables, independent of potential confounders and not subject to reverse causation. Using the epigenetic MR approach, our previous study on smoking-related DNA methylation and IBD has provided an exciting research paradigm to explore potential mechanisms of environmental effects in complex diseases by integrating observational, genetic and methylation evidence.⁹

Herein, we hypothesized that a potential pathogenic effect of air pollution on IBD risk is mediated by altered DNA methylation. To test this hypothesis, we first extended the study reported by Li et al.¹⁶ and tested these associations. Next, we further explored possible effects of genetic susceptibility and lifestyle factors, as well as the roles of deprivation and urbanization status as interaction factors in the association between air pollution and UC development. Then, we performed MR analysis to assess potential causal association between air pollution-related DNA methylation and UC risk. Genome-wide methylation study of IBD and gene expression on colonic tissues were also conducted to finally validate the above key findings.

Methods

Study design

The study design is displayed in Fig. 1.

Prospective cohort study

Study population

We used data from the UK Biobank, an ongoing national prospective cohort study with over 500,000 individuals in the UK recruited between 2006 and 2010.³⁴ We excluded participants 1) without available air pollution data; 2) with pre-diagnosed IBD at baseline or incident IBD in the first year of follow-up; 3) reporting moving after recruitment. A total of 452,012 (895 cases) and 453,199 (2082 cases) participants were eligible for CD and UC analysis, respectively.

Assessment of air pollution

The land use regression (LUR)-based model was used to estimate the annual average concentration of particulate

matter of fine inhalable particles, with diameters that are generally 2.5 µm and smaller (PM_{2.5}), particulate matter of inhalable particles, with diameters that are generally 10 µm and smaller (PM₁₀), nitrogen dioxide (NO₂), and nitrogen oxides (NO_x),³⁵ as previously described (Supplementary Table S1).³⁵ Previous studies compared and confirmed similar patterns of air pollution exposure in the UK Biobank cohort and the public UK Air Information Resources.^{36,37} Cumulative exposure to air pollution was calculated as pollutant concentration multiplied by the average daily time spent outdoors.³⁸ The combined air pollution exposure score was estimated as follows. Firstly, a multivariate Cox regression was applied to obtain the beta value of each pollutant, then the total effect of air pollution was developed by using the formula $[(\beta_{PM_{2.5}} * PM_{2.5} + \beta_{PM_{10}} * PM_{10} + \beta_{NO_2} * NO_2 + \beta_{NO_x} * NO_x) * (4/\text{sum of the } \beta \text{ coefficients})]$.³⁸ In calculation of each β value, taking $\beta_{PM_{2.5}}$ as an example, we first included participants' exposure to PM_{2.5} into the multivariable Cox proportional hazards model to estimate its effect on the risk of UC, and the partial regression coefficients for PM_{2.5} is $\beta_{PM_{2.5}}$.

Ascertainment of outcomes and covariates

The main endpoint of the study was the occurrence of incident IBD, including its subtypes UC and CD, throughout the follow-up period. Incident IBD were ascertained through data of primary care (converted into International Statistical Classification of Disease and Related Health Problems [ICD]-10 code), hospital inpatient (ICD-9 or ICD-10 code), and death registry (ICD-10 code). The ICD-9 and ICD-10 codes for UC were 556 and K51, while those for CD were 555 and K50. We excluded individuals with concurrent diagnoses of CD and UC when the outcomes of interest were specifically focused on either CD or UC, rather than IBD as a whole.

The following covariates collected at baseline were included in the analysis: age at recruitment, sex, ethnicity, education, household income, Townsend deprivation index (TDI), living area (urban or rural), alcohol consumption, smoking status, body mass index (BMI), physical activity, and IBD-associated healthy diet.³⁹ Consuming at least 4 of the following 7 food groups (≥ 3 servings fruits per day; ≥ 3 servings vegetables per day; ≥ 2 servings fish per week; ≤ 1 serving processed meats per week; ≤ 1.5 servings unprocessed red meats per week; ≥ 3 servings whole grains per day; ≤ 1.5 servings refined grains per day) indicated a healthy diet. Indoor air pollution, as indicated by the presence of smokers in the household and exposure to tobacco smoke at home, as well as the use of open gas or solid fuel fires for cooking/heating, was also taken into consideration in this analysis. We assessed genetic susceptibility to UC and CD for each participant through the development of polygenic risk scores (PRS). The construction of the PRS was based on 48 and 67 independent single-nucleotide polymorphisms (SNPs)

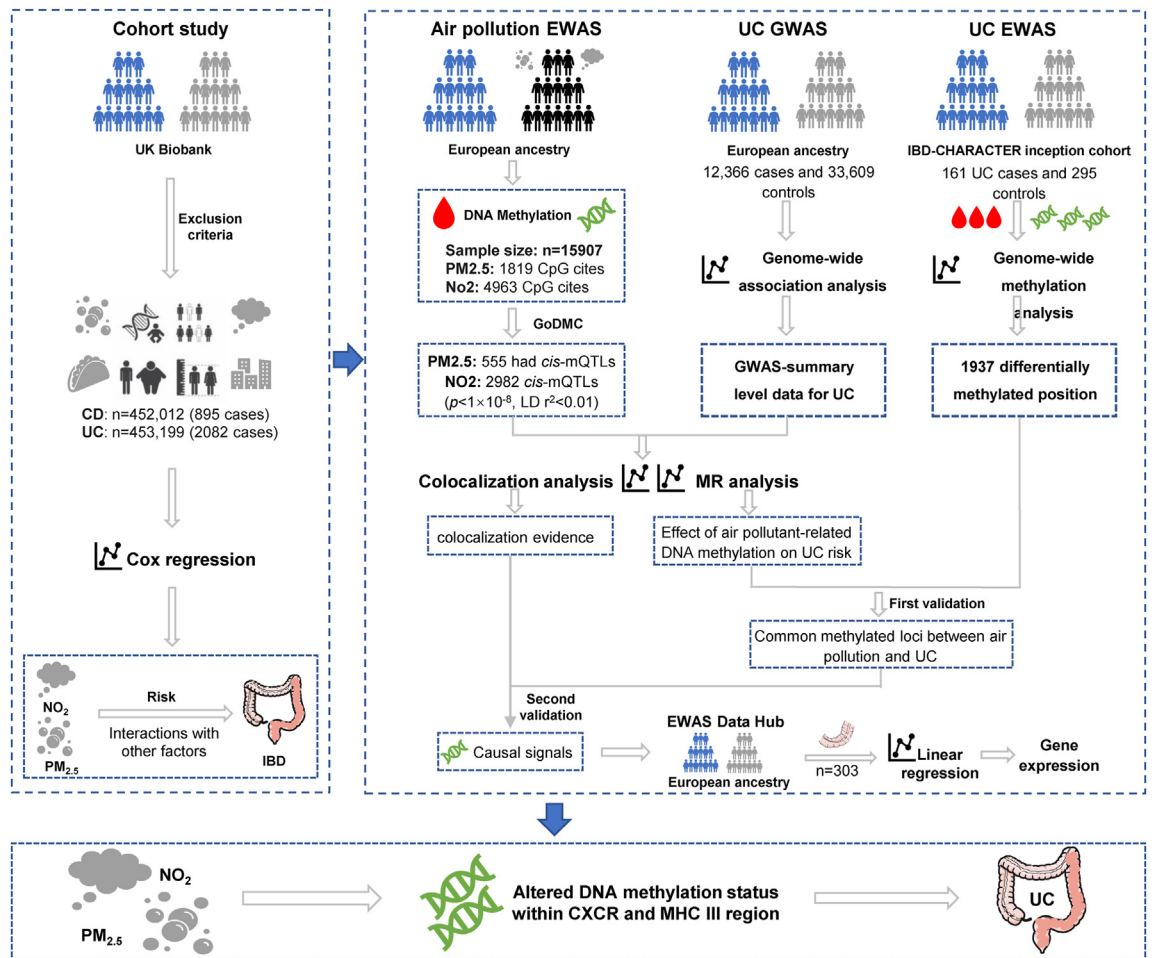


Fig. 1: Flow chart of the study. IBD, inflammatory bowel disease; CD, Crohn’s disease; UC, ulcerative colitis; PM, particulate matter; NO_x, nitrogen oxides; NO₂, nitrogen dioxide; EWAS, Epigenome-wide association studies; GWAS, genome-wide association study; MR, Mendelian randomization; mQTL, methylation quantitative trait loci; DNA, deoxyribonucleic acid; MHC, major histocompatibility complex; CXCR2, CXC motif chemokine receptor 2.

strongly associated with UC and CD ($P < 5 \times 10^{-8}$, $r^2 < 0.001$) (the lists and specific information of the SNPs have been described in detail in our previous study),⁶ respectively, identified from a European genome-wide association study (GWAS) involving 86,640 individuals (6968 UC and 5956 CD).⁴⁰ Then, the PRS was calculated by summing up the number of risk-increasing alleles for each SNP weighted by effect size on genetic liability, as described before.⁶ According to the PRS, the genetic risk of CD or UC in the population is divided into three levels: high (highest quintile), intermediate (quintiles 2–4) and low (the lowest quintile). Detailed information and calculation of the covariates are presented in [Supplementary Table S2](#).

Statistical analysis

We first conducted correlation analyses of urbanization and TDI with air pollution. The associations between air

pollution, cumulative exposure to air pollution, and combined air pollution exposure score (PM_{2.5}, PM₁₀, NO₂, NO_x, in quantiles, per interquartile range [IQR]) with the risk of IBD were estimated by performing Cox proportional hazard models. The basic model was adjusted for age, sex, and ethnicity, with the fully adjusted model further adjusted for physical activity, healthy diet, education, household income, smoking status, BMI, assessment centers, and alcohol consumption. Subsequently, the fully adjusted model was applied to evaluate the risk of UC and CD when exposed to air pollutants.

To explore the interactions between air pollution and genetics as well as lifestyle factors in the pathogenesis of UC, we conducted interaction analysis using both multiplicative and additive models.⁴¹ In subgroup analyses and Kaplan–Meier survival analysis, we examined differences in the risk of UC within categories of

different genetic risk (low, moderate, and high), smoking status (never, previous, or current), diet quality (healthy, unhealthy), BMI (<25 , ≥ 25 kg/m²), alcohol consumption (none to moderate, heavy), and physical activity (inadequate, adequate). In the joint analyses, participants were divided into several groups according to the combination of air pollution exposure score (lowest: 1st quintile; intermediate: 2nd–4th quintiles; high: 5th quintile) and genetic risk or the lifestyle factors, and taken individuals at low air pollution and low genetic risk (or never smoking; healthy diet; BMI <25 kg/m², none to moderate alcohol consumption, inadequate physical activity) as the reference group to evaluate the risk of UC of other groups. We also examined whether the primary findings were changed when additionally adjusted for PRS score or indoor air pollution related variables, and whether urbanization and TDI were linked to the development of IBD.

All analyses were conducted using R software, version 4.1.3. We corrected for multiple testing using a Bonferroni corrected *P*-value threshold ($0.05/5 = 0.01$).

Ethics

The UK Biobank received the ethical approval from the North West Multicenter Research Ethics Committee (REC reference: 21/NW/0157) and participants provided signed informed consent. The application number of this study is 73,595.

Epigenetic Mendelian randomization analysis

Genetic instruments of air pollution-related DNA methylation
The association estimates between PM_{2.5}⁴² and NO₂⁴³ and DNA methylation were derived from the corresponding genome-wide DNA methylation analyses. There are no available CpG sites or *cis*-mQTL for PM₁₀ and NO_x. Specifically, a total of 1819 and 4963 CpG sites were significantly differentially methylated when exposed to PM_{2.5} and NO₂, respectively (Supplementary Tables S3–S4). Next, methylation quantitative trait loci (mQTLs), which were robustly associated with the methylation of these CpG sites, were derived from the Genetics of DNA Methylation Consortium (GoDMC).⁴⁴ We extracted information on mQTLs including location, effect allele and other allele, and the corresponding effect estimates (including beta, se, and *P* value). Subsequently, the significant *cis*-mQTL ($P < 1 \times 10^{-8}$, the distance between mQTL and CpG site <1 MB) from the additive random-effects meta-analysis were extracted to proxy the methylation level for each CpG site associated with PM_{2.5} and NO₂ for UC analysis. Finally, we performed LD pruning ($r^2 > 0.01$) to filter independent instrumental variants for air pollution-related DNA methylation, and calculated the *F* statistic for each instrumental variable to exclude the weak instruments ($F_{\text{statistic}} < 10$). The detailed information on the instrumental variables are presented in Supplementary Tables S5–S6. To elucidate potential gene pathways,

methylation sites associated with air pollution were subjected to ontology analysis using the *misMethyl::gomethyl* function (v. 1.34.0).⁴⁵ This allows for enrichment analysis of a gene set and indicates which molecular functions and biological processes are over-represented. In the current ontology analysis, we input the significant CpGs, calculated the probability of genes being selected given the number of associated CpGs and tested for each gene ontology category.⁴⁶

Summary-level data of UC

Genetic associations with UC were obtained from the summary-level data of a large genome-wide association study (GWAS) conducted in European ancestry.⁴⁷ Briefly, de Lange et al. carried out a GWAS of 12,160 IBD cases and 13,145 controls and then meta-analyzed with previously published summary statistics from 12,882 IBD cases and 21,770 controls, bringing the sample size up to 45,975 participants (including 12,366 UC cases).⁴⁷

MR analysis

To investigate the potential causal associations between genetically determined air pollutants-related DNA methylation and UC risk, each CpG site was seen as an exposure, and its proxy mQTLs were used as instrumental variables. The Wald ratio approach was applied to estimate the effect of DNA methylation in each CpG site on per standard deviation (SD) increment of UC risk. If there were more than one independent mQTLs for each CpG site, we pooled the estimates using the random effects inverse-variance weighted (IVW) method. We also applied the MR-Egger approach to assess potential horizontal pleiotropy.⁴⁸ Bonferroni correction was adopted to correct for multiple testing. MR analyses were conducted in R software (version 4.1.3) by using the “TwoSampleMR” (version 0.5.6) R packages.

Genome-wide DNA methylation and co-localization analyses

To validate the loci identified in the epigenetic MR analysis, we conducted a genome-wide DNA methylation analysis for UC, including 161 cases and 295 controls.²⁷ The detailed information and procedures of genome-wide DNA methylation profiling for UC have been described elsewhere.²⁷ In this analysis, we constructed a specific hypothesis based on the findings from discovery stage; as such a *P*-value threshold (<0.05) was therefore deemed acceptable and adopted for corroboration. In order to validate the proposed associations between loci exhibiting altered DNA methylation in response to air pollution and the risk of UC, we conducted a comparative analysis between the findings from epigenetic MR analysis and genome-wide methylation analysis of UC. This allowed us to identify overlapping significant gene loci that are implicated in

both air pollution-induced methylation changes and UC-related methylation alterations.

We additionally conducted a co-localization analysis to explore whether the associations between the methylation level of specific CpG sites and UC risk were driven by a shared causal variant. Specifically, we extracted all available mQTLs for every significant CpG site from GoDMC and integrated them with GWAS summary-level data of UC. A summary posterior probability for the CpG site and a posterior probability for the single mQTL used as a genetic IV at 80% or higher were considered evidence of colocalization. The “coloc” R package in R software (version 4.1.3) was used to perform this analysis.⁴⁹

DNA methylation and gene expression, single cell RNA sequencing analysis in intestinal tissues

For those CpG sites that were validated in the genome-wide methylation analysis or were proved to have convincing evidence of colocalization, we performed linear regression to explore the regulation pattern of DNA methylation on gene expression in colonic tissue by using data provided by the EWAS Data Hub (<https://ngdc.cncb.ac.cn/ewas/datahub/index>). The EWAS Data Hub is a resource used to collect and normalize DNA methylation array data and archived associated data, including 75,344 samples to date, and providing reference DNA methylation profiles under different contexts, involving 81 tissues or cell types, 6 ancestry categories, and 67 diseases.^{50,51}

In order to delineate the specific distribution patterns of target genes across cell subtypes, we conducted an analysis of their expression levels in RNA-seq datasets GSE150516 and GSE231993 utilizing the Seurat V5 package. These datasets comprised 8 inflamed tissues from patients with UC, 9 uninflamed tissues from self-control patients with UC, and 4 normal tissues from healthy colon samples. Cells were filtered based on the following criteria: 1) nFeature_RNA >300 and nFeature_RNA <70,000; 2) nCount_RNA <100,000; 3) percentage of mitochondria <20%; 4) percentage of red blood cells <3%. Doublet cells were identified and removed using the DoubletFinder package. Further analysis included batch effect correction using the ‘IntegrateLayers’ function with the ‘harmony’ algorithm, after normalizing and scaling UMI counts and identifying highly variable features. Integrated cells were then dimensionally reduced using the Uniform Manifold Approximation and Projection (UMAP) algorithm with a resolution of 0.5. Cell types were annotated using the SingleR package with known markers (‘ENG’, ‘PECAM1’ for endothelial cells; ‘COL14A1’, ‘COL3A1’, ‘DCN’, ‘SPARC’ for fibroblasts; ‘PHGR1’, ‘TFF3’, ‘ELF3’, ‘KRT19’, ‘CLDN3’, ‘CLDN7’ for epithelial cells; ‘CD163’, ‘CXCL8’, ‘AIF1’, ‘FCER1G’, ‘CPA3’ for myeloid cells; ‘CD79A’, ‘CD79B’, ‘MS4A1’, ‘VPREB3’ for B cells; ‘GZMA’, ‘GZMB’, ‘GZMK’, ‘PRF1’, ‘CCL5’

for NK cells; ‘CD2’, ‘IL7R’, ‘CD3E’, ‘CD3G’, ‘TRBC2’ for T cells; ‘JCHAIN’, ‘MZB1’, ‘IGHA1’, ‘DERL3’ for plasma cells). Finally, we compared the expression levels and percentages of target genes across different cell subtypes between inflamed UC tissues and normal colon tissues, and assessed significance using the Wilcoxon method.

Role of funders

The funding sources had no role in the design of this study and did not have any role in data collection, data analyses, interpretation, writing of report, or decision to submit results.

Results

UK Biobank-based cohort study: air pollution was associated with incident UC but not CD

With a mean (SD) follow-up of 13.2 (2.1) years, 2082 incident UC and 895 incident CD were documented in the UK biobank cohort ([Supplementary Table S7](#)). The mean age of overall participants, individuals with UC and CD is 56.5, 57.5 and 56.9 years old. NO_x, NO₂, PM₁₀, PM_{2.5} and the combined air pollution exposure score showed a correlation with higher urbanization ($r = 0.45, 0.52, 0.37, 0.45, 0.50$) and Townsend deprivation index ($r = 0.44, 0.44, 0.37, 0.42, 0.47$) ([Supplementary Table S8](#)). Compared with the lowest quantile of exposure, higher quantile exposure of NO_x, NO₂, PM_{2.5} and a combined air pollution exposure score were associated with increased risk of UC, with HR (95% CI) being 1.20 (95% CI: 1.05–1.38), 1.19 (95% CI: 1.03–1.36), 1.19 (95% CI: 1.05–1.36), 1.26 (95% CI: 1.11, 1.45) ([Table 1](#)). These associations remained consistent in all sensitivity analyses ([Supplementary Table S9](#)). In addition, the observed associations with UC remained significant when considering the cumulative exposure to air pollutants ([Supplementary Table S10](#)). No significant association between air pollutants and CD risk was observed in this prospective cohort study.

Interactions between effects of air pollution, genetic susceptibility, and other lifestyle factors

In our stratification analysis, we observed that the associations between exposure to intermediate and high levels of air pollution and the risk of UC showed higher HR ratio estimates in the subgroup of participants with a high genetic risk, an unhealthy diet, a history of smoking, heavy alcohol consumption, and a higher BMI (≥ 25 kg/m²) ([Table 2](#)), in line with the association trends showed by the Kaplan–Meier plots. As shown in [Supplementary Figure S1](#), exposure to low, intermediate and high air pollution levels showed different risk of UC in subgroup participants stratified by diet, smoking status, alcohol consumption status and physical activity (P log-rank <0.05), although not significant in participants with low genetic risk (P log-rank = 0.83), lower

	IBD				P	CD	P	
	Model 1	P	Model 2	P				
	HR (95% CI) ^a		HR (95% CI) ^b					
NO_x								
Per IQR	1.07 (1.03, 1.11)	1.23E-04	1.04 (1.00, 1.08)	4.23E-02	1.06 (1.02, 1.11)	4.04E-03	0.97 (0.90, 1.05)	4.34E-01
Q1	Ref		Ref		Ref		Ref	
Q2	1.02 (0.91, 1.14)	7.84E-01	0.98 (0.87, 1.10)	6.88E-01	1.02 (0.89, 1.17)	7.63E-01	0.89 (0.72, 1.09)	2.53E-01
Q3	1.16 (1.04, 1.30)	6.65E-03	1.10 (0.98, 1.23)	9.80E-02	1.19 (1.04, 1.36)	1.13E-02	0.91 (0.74, 1.12)	3.65E-01
Q4	1.04 (0.93, 1.16)	4.83E-01	0.96 (0.86, 1.08)	5.06E-01	1.07 (0.93, 1.22)	3.53E-01	0.80 (0.64, 1.01)	5.66E-02
Q5	1.22 (1.09, 1.36)	3.99E-04	1.13 (1.01, 1.26)	3.87E-02	1.20 (1.05, 1.38)	6.89E-03	0.96 (0.78, 1.18)	6.95E-01
P-trend		7.40E-04		8.04E-02		7.18E-03		2.96E-01
NO₂								
Per IQR	1.08 (1.04, 1.12)	1.66E-04	1.06 (1.01, 1.10)	9.78E-03	1.07 (1.02, 1.12)	5.55E-03	1.02 (0.94, 1.10)	6.78E-01
Q1	Ref		Ref		Ref		Ref	
Q2	1.09 (0.97, 1.21)	1.49E-01	1.04 (0.92, 1.16)	5.51E-01	1.06 (0.93, 1.22)	3.91E-01	0.97 (0.79, 1.21)	8.06E-01
Q3	1.14 (1.02, 1.27)	1.97E-02	1.08 (0.96, 1.20)	2.12E-01	1.10 (0.96, 1.25)	1.84E-01	1.03 (0.83, 1.27)	8.19E-01
Q4	1.18 (1.05, 1.31)	3.97E-03	1.09 (0.97, 1.22)	1.52E-01	1.13 (0.98, 1.29)	8.24E-02	0.99 (0.80, 1.23)	9.34E-01
Q5	1.22 (1.09, 1.36)	6.08E-04	1.14 (1.02, 1.28)	2.55E-02	1.19 (1.03, 1.36)	1.47E-02	1.03 (0.83, 1.29)	7.65E-01
P-trend		2.14E-04		1.81E-02		1.02E-02		7.39E-01
PM₁₀								
Per IQR	1.03 (0.99, 1.07)	1.87E-01	1.01 (0.97, 1.06)	5.93E-01	1.02 (0.97, 1.08)	4.16E-01	0.99 (0.91, 1.07)	7.56E-01
Q1	Ref		Ref		Ref		Ref	
Q2	1.10 (0.98, 1.22)	9.44E-02	1.09 (0.97, 1.21)	1.47E-01	1.14 (1.00, 1.29)	5.30E-02	0.96 (0.77, 1.19)	6.89E-01
Q3	1.05 (0.94, 1.17)	3.77E-01	1.04 (0.93, 1.16)	5.41E-01	1.05 (0.92, 1.19)	5.04E-01	1.01 (0.82, 1.24)	9.37E-01
Q4	0.94 (0.84, 1.05)	2.87E-01	0.90 (0.80, 1.01)	7.40E-02	0.88 (0.76, 1.00)	5.85E-02	0.96 (0.78, 1.19)	7.07E-01
Q5	1.09 (0.97, 1.22)	1.33E-01	1.06 (0.95, 1.20)	2.86E-01	1.09 (0.95, 1.25)	2.05E-01	1.00 (0.80, 1.24)	9.75E-01
P-trend		9.16E-01		5.88E-01		5.33E-01		9.76E-01
PM_{2.5}								
Per IQR	1.10 (1.05, 1.14)	1.09E-05	1.06 (1.02, 1.11)	7.92E-03	1.08 (1.02, 1.13)	4.58E-03	1.02 (0.94, 1.11)	6.44E-01
Q1	Ref		Ref		Ref		Ref	
Q2	1.02 (0.92, 1.15)	6.78E-01	0.98 (0.88, 1.10)	7.80E-01	1.02 (0.89, 1.17)	8.13E-01	0.91 (0.73, 1.12)	3.71E-01
Q3	1.07 (0.96, 1.20)	2.27E-01	1.01 (0.90, 1.13)	8.52E-01	1.06 (0.92, 1.21)	4.10E-01	0.90 (0.73, 1.12)	3.37E-01
Q4	1.18 (1.06, 1.31)	3.06E-03	1.11 (0.99, 1.24)	6.39E-02	1.19 (1.05, 1.36)	8.72E-03	0.92 (0.75, 1.14)	4.63E-01
Q5	1.20 (1.07, 1.33)	1.31E-03	1.10 (0.98, 1.23)	9.94E-02	1.13 (0.99, 1.29)	7.30E-02	1.03 (0.83, 1.27)	8.05E-01
P-trend		6.09E-05		1.49E-02		7.52E-03		7.61E-01
Combined air pollution exposure score								
Per IQR	1.09 (1.05, 1.14)	1.75E-05	1.06 (1.01, 1.10)	9.04E-03	1.08 (1.03, 1.13)	2.58E-03	1.00 (0.92, 1.09)	9.36E-01
Q1	Ref		Ref		Ref		Ref	
Q2	1.08 (0.97, 1.21)	1.60E-01	1.03 (0.92, 1.15)	6.51E-01	1.07 (0.93, 1.22)	3.49E-01	0.93 (0.76, 1.16)	5.33E-01
Q3	1.14 (1.02, 1.28)	1.90E-02	1.06 (0.95, 1.19)	2.88E-01	1.12 (0.98, 1.28)	9.74E-02	0.93 (0.76, 1.15)	5.25E-01
Q4	1.11 (0.99, 1.24)	7.57E-02	1.02 (0.91, 1.14)	7.49E-01	1.10 (0.96, 1.26)	1.68E-01	0.84 (0.67, 1.04)	1.08E-01
Q5	1.29 (1.16, 1.44)	4.98E-06	1.20 (1.07, 1.34)	1.89E-03	1.26 (1.11, 1.45)	6.46E-04	1.05 (0.85, 1.29)	6.79E-01
P-trend		2.04E-05		6.39E-03		1.15E-03		9.36E-01

IBD, inflammatory bowel disease; UC, ulcerative colitis; CD, Crohn's disease; CI, confidence interval; HR, hazard ratio; IQR, interquartile range; NO₂, nitrogen dioxide; NO_x, nitrogen oxides; PM, particulate matter. *P* < 0.05 and >0.05/5 (0.01) was considered suggestive, *P* < 0.05/5 (0.01) after Bonferroni correction was considered significant and was bolded. ^aModel 1 was adjusted for age, sex, ethnicity. ^bModel 2 was adjusted for age, sex, ethnicity, assessment centers, household income, smoking status, education, BMI, physical activity, healthy diet, and alcohol consumption.

Table 1: Associations of air pollutants with risk of incident IBD.

BMI (<25 kg/m²) (*P* log-rank = 0.34). Suggestive multiplicative interactions were observed for air pollution with diet (*P*_{interaction} = 4.00E-02, likelihood ratio test) and alcohol consumption (*P*_{interaction} = 2.79E-03, likelihood ratio test), and significant additive interactions were only

observed for air pollution with genetic risk (relative excess risk due to interaction [RERI] = 0.95, 95% CI: 0.34–1.57; *P*_{interaction} = 1.23E-03, likelihood ratio test) (Table 2). In joint analyses of air pollution score with genetic risk, the HR of UC showed a linear increase

	Subgroup analysis			Additive interaction	
	HR (95% CI)	P	P-interaction for multiplicative model	RERI (95% CI)	P-interaction for additive model
Genetic risk					
Low genetic risk	Ref		2.75E-01	0.95 (0.34, 1.57)	1.23E-03
Low air pollution level	1.02 (0.76, 1.38)	8.87E-01			
Intermediate air pollution level	0.97 (0.65, 1.44)	8.75E-01			
High air pollution level					
Intermediate genetic risk	Ref				
Low air pollution level	1.13 (0.97, 1.33)	1.22E-01			
Intermediate air pollution level	1.29 (1.06, 1.58)	1.16E-02			
High air pollution level					
High genetic risk	Ref				
Low air pollution level					
Intermediate air pollution level	1.28 (1.00, 1.63)	4.78E-02			
High air pollution level	1.64 (1.23, 2.20)	7.92E-04			
Diet					
Healthy diet			4.00E-02	0.21 (-0.12, 0.53)	1.06E-01
Low air pollution level	Ref				
Intermediate air pollution level	1.00 (0.87, 1.14)	9.81E-01			
High air pollution level	1.22 (1.04, 1.43)	1.65E-02			
Unhealthy diet					
Low air pollution level	Ref				
Intermediate air pollution level	1.37 (1.10, 1.70)	4.47E-03			
High air pollution level	1.42 (1.10, 1.84)	6.48E-03			
BMI					
BMI <25			1.01E-01	0.31 (0.04, 0.59)	1.25E-02
Low air pollution level	Ref				
Intermediate air pollution level	0.91 (0.75, 1.11)	3.47E-01			
High air pollution level	1.02 (0.80, 1.30)	8.68E-01			
BMI ≥25					
Low air pollution level	Ref				
Intermediate air pollution level	1.20 (1.05, 1.38)	9.70E-03			
High air pollution level	1.41 (1.19, 1.66)	4.96E-05			
Smoking status					
Never			2.10E-01	0.22 (-0.10, 0.55)	8.89E-02
Low air pollution level	Ref				
Intermediate air pollution level	0.99 (0.84, 1.16)	8.77E-01			
High air pollution level	1.21 (0.99, 1.48)	6.56E-02			
Ever smoked					
Low air pollution level	Ref				
Intermediate air pollution level	1.20 (1.03, 1.41)	2.17E-02			
High air pollution level	1.32 (1.10, 1.59)	2.74E-03			
Alcohol consumption					
None to moderate			2.79E-03	0.01 (-0.30, 0.32)	4.76E-01
Low air pollution level	Ref				
Intermediate air pollution level	1.02 (0.90, 1.16)	7.61E-01			
High air pollution level	1.27 (1.10, 1.48)	1.64E-03			
Heavy					
Low air pollution level	Ref				
Intermediate air pollution level	1.45 (1.12, 1.88)	5.01E-03			
High air pollution level	1.26 (0.92, 1.72)	1.46E-01			

(Table 2 continues on next page)

	Subgroup analysis			Additive interaction	
	HR (95% CI)	P	P-interaction for multiplicative model	RERI (95% CI)	P-interaction for additive model
(Continued from previous page)					
Physical activity					
Inadequate			9.41E-01	0.05 (-0.28, 0.38)	3.83E-01
Low air pollution level	Ref				
Intermediate air pollution level	1.11 (0.91, 1.36)	3.06E-01			
High air pollution level	1.26 (0.99, 1.60)	5.60E-02			
Adequate					
Low air pollution level	Ref				
Intermediate air pollution level	1.09 (0.95, 1.25)	2.13E-01			
High air pollution level	1.27 (1.08, 1.49)	4.76E-03			

HR and 95% CI was calculated by Cox model adjusted for age, sex, ethnicity, assessment centers, household income, smoking status, education, BMI, physical activity, healthy diet, and alcohol consumption. Given that for all genetic-related analysis, only participants of White genetic background were included, the analysis was not adjusted for ethnicity. BMI, body mass index; HR, hazard ratio; CI, confidence interval; RERI, relative excess risk due to interaction. A RERI of greater than zero means positive interaction or more than additivity. A RERI of less than zero means negative interaction or less than additivity. Healthy diet indicated satisfying at least 4 of the following 7 food consumptions (≥ 3 servings fruits per day; ≥ 3 servings vegetables per day; ≥ 2 servings fish per week; ≤ 1 serving processed meats per week; ≤ 1.5 servings unprocessed red meats per week; ≥ 3 servings whole grains per day; ≤ 1.5 servings refined grains per day), and unhealthy diet indicated satisfying less than 4 of the following 7 food consumptions. High, intermediate, and low genetic risk indicated the highest quintile, quintiles 2–4 and lowest quintile of polygenic risk score. $P < 0.05$ and $>0.05/4$ (0.0125) was considered suggestive, $P < 0.05/4$ (0.0125) after Bonferroni correction was considered significant and was bolded.

Table 2: Interaction analyses of air pollution levels with genetic risk and lifestyle factors on the risk of incident ulcerative colitis.

with the increasing levels of air pollution. Compared with individuals with low genetic risk and low air pollution level, the HRs of UC for those with high genetic risk and high air pollution level was 2.34 (95% CI: 1.70–3.23). Similarly, compared with low air pollution and healthy lifestyle (never smoking; healthy diet; BMI < 25 ; inadequate physical activity), exposure to high levels of air pollution and unhealthy lifestyle was associated with increased risk of UC (Table 3).

In our analysis of the impact of deprivation status and urban/rural location on the risk of IBD, we observed a significant association between higher TDI and an increased risk of UC as well as CD (UC: HR_{highest tertile} = 1.21, 95% CI 1.10–1.35; CD: HR_{highest tertile} = 1.35, 95% CI 1.14–1.60). However, we did not find a significant association between living in an urban area and UC nor CD (UC: HR = 1.06, 95% CI 0.94–1.19; CD: HR = 1.08, 95% CI 0.89–1.31). The association was no longer significant when further adjusted for air pollution and smoking status (HR = 1.10, 95% CI 0.98–1.24) (Supplementary Table S11).

Blood DNA methylation alterations of air pollutants and UC risk

Subsequent analyses were exclusively focused on UC due to the significance of correlation effects observed between air pollution and UC, as opposed to CD. Among the 1819 CpG sites that were reported to be associated with PM_{2.5}, 555 had available *cis*-mQTLs that could be applied as proxies in the two-sample MR analysis. We found that altered methylation at five PM_{2.5}-related CpG sites was significantly associated with

UC risk ($P_{\text{Bonferroni}} = 0.05/555$, 9.00E-05), including cg16689962 (*AGPAT1*; odds ratio [OR] = 9.77, 95% CI: 6.45–14.80), cg01710852 (*SMAD3*; OR = 1.88, 95% CI: 1.49–2.38), cg11359771 (*P4HA2*; OR = 1.83, 95% CI: 1.44–2.33), cg24143221 (*HSPA1L*; OR = 0.58, 95% CI: 0.45–0.74), and cg27490128 (*NCR3*; OR = 0.41, 95% CI: 0.26–0.63) (Fig. 2, Supplementary Table S12). There was no horizontal pleiotropy for the used instruments for PM_{2.5}-related DNA methylation (Supplementary Table S13).

For NO₂, 2928 of the 4963 CpG sites were discovered to have *cis*-mQTLs that could proxy NO₂-related DNA methylation in the two-sample MR analysis (Supplementary Table S14). We observed that altered methylation at 22 NO₂-related CpG sites was significantly related to UC risk ($P_{\text{Bonferroni}} = 0.05/2928$, 1.70E-05), including cg24011261 (*GPX1*), cg10502563 (*EGFL8*), cg22250546 (*CEBPA*), cg25953682 (*DDAH2*), cg00446123 (*LIME1*), and cg06547715 (*CXCR2*) (Fig. 3, Supplementary Table S15). Horizontal pleiotropy was observed in the instrumental variants for cg13012653 (*CAST*) and cg15818109 (*COL11A2*) (Supplementary Table S15).

In the ontology analysis of air pollution-related methylation sites, CpGs related to NO₂ were enriched for terms pointing at the role of secretory vesicles (false discovery rate [FDR] = 2.99E-11), the lysosome (FDR = 2.30E-3), and leukocyte activation (FDR = 0.018), adaptive immunity (FDR = 2.00E-5), and T-cell receptor signaling (FDR = 4.27E-4) (Supplementary Table S16). CpG sites related to PM_{2.5} were not enriched for any biological processes.

	Incident UC	
	HR (95% CI)	P
Genetic risk		
Low genetic risk		
Low air pollution level	Ref	
Intermediate air pollution level	1.03 (0.77, 1.39)	8.28E-01
High air pollution level	0.96 (0.65, 1.43)	8.50E-01
Intermediate genetic risk		
Low air pollution level	1.14 (0.85, 1.52)	3.89E-01
Intermediate air pollution level	1.30 (0.99, 1.70)	5.71E-02
High air pollution level	1.47 (1.09, 1.97)	1.04E-02
High genetic risk		
Low air pollution level	1.41 (1.01, 1.97)	4.38E-02
Intermediate air pollution level	1.77 (1.34, 2.35)	6.86E-05
High air pollution level	2.34 (1.70, 3.23)	2.27E-07
Smoking status		
Never		
Low air pollution level	Ref	
Intermediate air pollution level	0.99 (0.84, 1.16)	8.74E-01
High air pollution level	1.19 (0.98, 1.45)	8.54E-02
Ever smoked		
Low air pollution level	1.30 (1.06, 1.58)	1.05E-02
Intermediate air pollution level	1.56 (1.33, 1.83)	3.25E-08
High air pollution level	1.74 (1.45, 2.09)	2.40E-09
Diet		
Healthy diet		
Low air pollution level	Ref	
Intermediate air pollution level	1.00 (0.88, 1.14)	9.91E-01
High air pollution level	1.21 (1.03, 1.42)	1.87E-02
Unhealthy diet		
Low air pollution level	0.93 (0.74, 1.17)	5.47E-01
Intermediate air pollution level	1.27 (1.10, 1.48)	1.46E-03
High air pollution level	1.34 (1.10, 1.63)	3.63E-03
Body mass index		
BMI <25		
Low air pollution level	Ref	
Intermediate air pollution level	0.92 (0.76, 1.13)	4.30E-01
High air pollution level	1.08 (0.85, 1.36)	5.47E-01
BMI ≥25		
Low air pollution level	0.87 (0.70, 1.08)	2.14E-01
Intermediate air pollution level	1.04 (0.85, 1.26)	7.15E-01
High air pollution level	1.19 (0.96, 1.48)	1.07E-01
Alcohol consumption		
None to moderate		
Low air pollution level	Ref	
Intermediate air pollution level	1.01 (0.89, 1.15)	8.71E-01
High air pollution level	1.25 (1.08, 1.46)	3.13E-03
Heavy		
Low air pollution level	0.72 (0.56, 0.93)	1.29E-02
Intermediate air pollution level	1.09 (0.93, 1.28)	2.98E-01
High air pollution level	0.96 (0.76, 1.21)	7.38E-01
Physical activity		
Inadequate		
Low air pollution level	Ref	
Intermediate air pollution level	1.09 (0.95, 1.25)	2.01E-01

(Table 3 continues on next page)

Validation and colocalization of differentially methylated genetic signals

When comparing the epigenetic MR findings with the genome-wide DNA methylation analysis of UC (Supplementary Table S17), we successfully validated the association of DNA methylation at loci including *AGPAT1* and *DDAH2* in MHC class III regions and *CXCR2* with the risk of UC (Table 4). As shown in Fig. 4 and Supplementary Table S18, we found that NO₂-related DNA methylation at cg06547715 [*CXCR2*] had a 96.9% posterior probability of sharing a causal variant (rs4133195) with UC susceptibility. However, due to there being less than 10 mQTLs for cg16689962, we could not perform a colocalization analysis to gain evidence supporting its causal effect on UC susceptibility.

The regulation pattern of DNA methylation on the target gene expression in colon tissue and the single cell RNA sequencing analysis

The altered DNA methylation at cg06547715 showed an effect on the expression of *CXCR2* in colonic tissues, with an effective coefficient of 1.60 (se = 0.66). In addition, the expression of genes especially in MHC class III regions was significantly regulated by the altered DNA methylation at cg16689962 (*AGPAT1*: beta = 0.65; *TNXB*: beta = 2.65; *TNF*: beta = 1.67). The scatter plots for the DNA methylation and gene expression profiles are presented in Supplementary Figure S2. To investigate the potential influence of mQTLs on the expression of their corresponding genes in intestinal tissue, we conducted a thorough inquiry using the GTEx Portal and revealed that the mQTL rs3132934 of cg16689962 serves as an eQTL for the *AGPAT1* gene in sigmoid and transverse colon tissue, with respective m-values of 0.972 and 0.981 (Supplementary Figure S3). Additionally, we performed a single-cell RNA sequencing analysis on intestinal tissue to assess the expression of these two genes across various intestinal cell types. Nine distinct cell types were identified and annotated within the intestinal tissue (Supplementary Figure S4a). The expression of *AGPAT1* and *CXCR2* was illustrated in Supplementary Figure S4b–d. It was observed that the expression of these two genes was significantly elevated in epithelial cells of patients with UC (Supplementary Figure S4c–f).

Discussion

The current study, based on data from observational and genetic studies, defines potentially causal associations between exposure to air pollution and the development of UC and illustrates the mechanistic importance of DNA methylation in the pathophysiological pathway. Specifically, we report an observational association between air pollution and the risk of incident UC among 453,919 individuals, which are primarily driven by associations between NO_x, NO₂, PM_{2.5} and UC. We also

report findings for the interactions between air pollution and genetic risk and specific lifestyle factors in the pathogenesis of IBD. Higher UC risk estimates were observed among groups with higher exposure to air pollution and higher genetic risk or unhealthy lifestyle (ever smoked, unhealthy diet, heavy alcohol consumption, BMI ≥ 25 kg/m²) in both subgroup and joint analyses. Subsequent MR-based analysis yielded potential causal evidence for the association between air pollution-related methylation and the risk of UC, identifying five PM_{2.5} and 22 NO₂-related CpG sites as possible contributors. Further genome-wide methylation analysis of UC, colocalization analysis, and gene expression and single cell RNA sequencing analyses in intestinal tissues provided a hierarchy of evidence to implicate the epigenetic alterations in *CXCR2* and loci in MHC class III region on the association of air pollutants with UC. Overall, our study provides a comprehensive assessment and exploration of the association and potential mechanisms linking air pollution to UC.

Previous evidence regarding the role of air pollution in IBD pathogenesis from case-control studies conducted in the Health Improvement Network (THIN)¹⁵ and the European Prospective Investigation into Cancer and Nutrition cohort (EPIC)¹⁴ did not report any significant associations. This may be due to the case-control design and relatively small number of IBD cases (367 CD and 591 UC). In contrast, a Canada population-based cohort study revealed an association between O_x and increased risk of pediatric-onset IBD,¹² and a modelling study found that PM_{2.5} environmental exposure was associated with pediatric IBD.⁵² The most recent study involving 450,000 participants, with more individuals with UC than CD, conducted in UK Biobank proposed associations between exposure to PM_{2.5}, PM₁₀, NO₂, and NO_x and the risk of UC.¹⁶ Our research was prompted by and built upon this index study; we expanded upon the findings with updated data and further assessed the interactions between genetic factors, lifestyle variables, and air pollution through subgroup, interaction, and joint analyses. Additionally, we examined the impact of urbanization and deprivation. We confirmed that high levels of air pollutants are associated with increased risk of UC, and the new findings of interactions between air pollution and genetics as well as lifestyle in the development of UC. Significant interactions were observed for genetic risk, alcohol consumption, and suggestive interactions were discovered for diet and BMI in the associations of air pollutants with UC risk. This indicates that genetic risk and alcohol consumption play a more important role in the observed effects compared to other lifestyle factors. High genetic risk, an unhealthy diet, higher BMI and a personal history of smoking were implicated as effect modifiers of the association between air pollution and incident UC. Of note, participants with a high genetic risk and high air pollution exposure showed higher risk

	Incident UC	
	HR (95% CI)	P
(Continued from previous page)		
High air pollution level	1.28 (1.09, 1.50)	3.23E-03
Adequate		
Low air pollution level	1.16 (0.94, 1.44)	1.74E-01
Intermediate air pollution level	1.28 (1.10, 1.49)	1.39E-03
High air pollution level	1.44 (1.19, 1.75)	2.37E-04

BMI, body mass index; UC, ulcerative colitis; HR, hazard ratio; CI, confidence interval. HR and 95% CI was calculated by Cox model adjusted for age, sex, ethnicity, assessment centers, household income, smoking status, education, BMI, physical activity, healthy diet, and alcohol consumption. Given that for all genetic-related analysis, only participants of White genetic background were included, the analysis was not adjusted for ethnicity. Healthy diet indicated satisfying at least 4 of the following 7 food consumptions (≥ 3 servings fruits per day; ≥ 3 servings vegetables per day; ≥ 2 servings fish per week; ≤ 1 serving processed meats per week; ≤ 1.5 servings unprocessed red meats per week; ≥ 3 servings whole grains per day; ≤ 1.5 servings refined grains per day), and unhealthy diet indicated satisfying less than 4 of the following 7 food consumptions. High, intermediate, and low genetic risk indicated the highest quintile, quintiles 2-4 and lowest quintile of polygenic risk score. $P < 0.05$ and $> 0.05/4$ (0.0125) was considered suggestive, $P < 0.05/4$ (0.0125) after Bonferroni correction was considered significant.

Table 3: Joint effects of combined air pollution levels, genetic risk and lifestyle factors for the risk of ulcerative colitis.

estimates than those with lowest genetic risk but high air pollution, implicating the importance of genetics in shaping the risk of UC when exposed to air pollution. Thus, the interaction and joint analyses explored how genetic and lifestyle play roles on air pollution and UC risk from different dimensions and presented preliminary evidence.

Furthermore, we demonstrate significant correlations between air pollution and urbanization, as well as between pollution and deprivation assessed by the TDI. This is consistent with the previously conducted study, which reported strong positive associations for air pollutants with urbanization and socioeconomic deprivation.⁵³ The observation that the positive correlations between deprivation and the risk of UC diminish in statistical significance upon further adjustment for both air pollution and smoking status is noteworthy, suggesting that these two factors may serve as pivotal determinants within deprived areas. These findings reveal the complex associations between deprivation, air pollution and IBD. Our earlier studies of children in Scotland have highlighted this complexity as well, demonstrating CD but not UC to be associated with affluence, rather than deprivation.⁵⁴ Further research is warranted to explore innovative potential avenues for investigating the relationship between the environment and the development of IBD. While our observations have been conducted in the UK, these findings may have direct implications for the rapid surge in IBD cases in newly industrialized nations.

The absence of any significant associations between air pollutants and the risk of developing CD is particularly noteworthy. Combining our findings with previous studies, we believe that there are several possible explanations. These include both statistical explanations

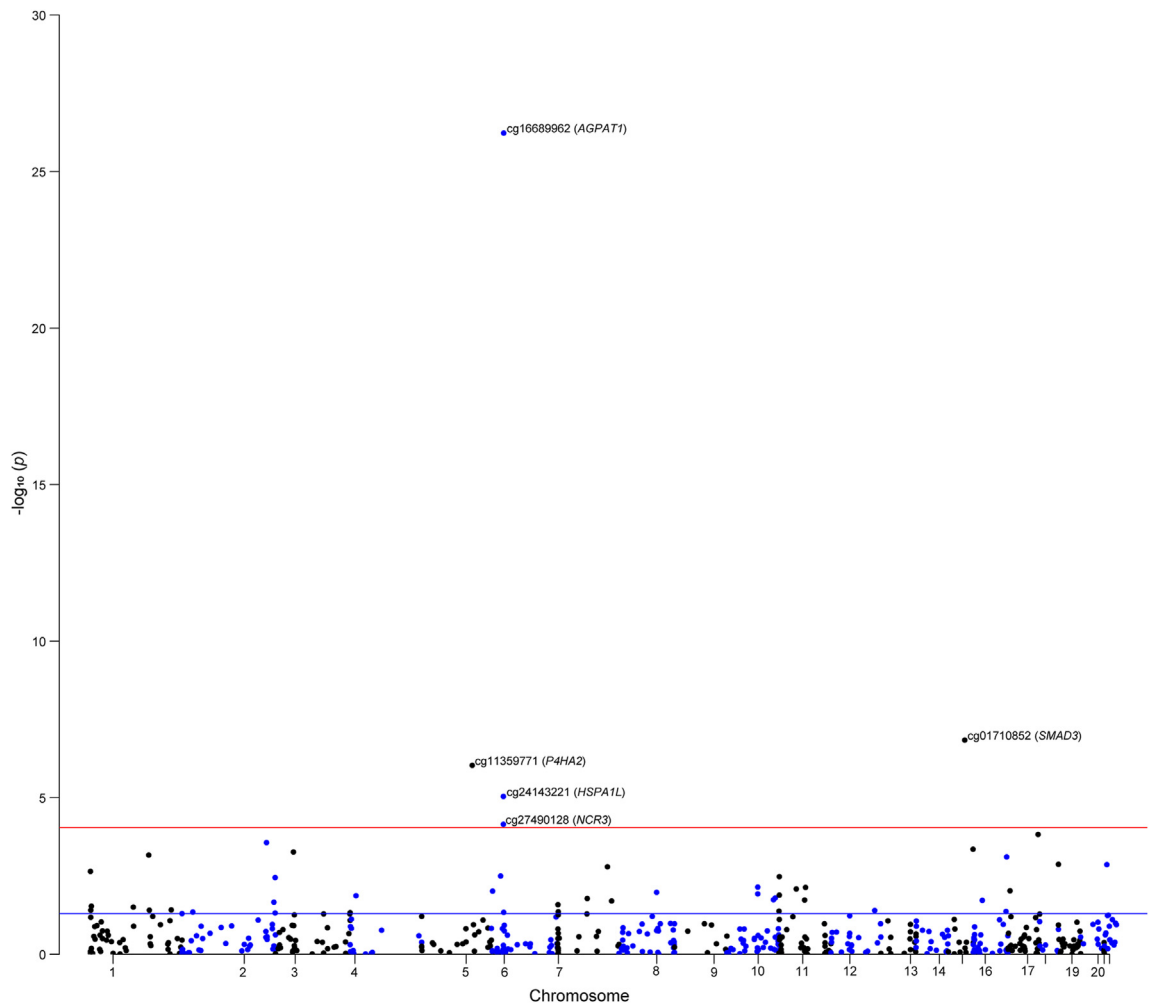


Fig. 2: The effect of gene regulated by $PM_{2.5}$ -related DNA methylation on the risk of ulcerative colitis. The blue line means $P = 0.05$, and the red line means the threshold of FDR correction.

based on power as well as biological explanations. Data suggest CD & UC are related to polygenic diseases—these share some genetic determinants, with others specific for each disease. We suggest a similar pattern is emerging for the exposome. The degree to which environmental risk factors are shared between CD and UC has been explored in epidemiological studies and meta-analyses.^{7,55} In our recent studies, we explored effects of a series of lifestyle factors on the development of CD and UC, respectively, and have established the positive associations of current smoking and previous smoking,⁹ unfavourable lifestyles (ever smoking, obesity, sleep <7 or >8 h/day, unhealthy dietary habits, unregular physical activity),⁶ short sleep duration and daytime napping⁵⁶ with higher risk of the both IBD subtypes, and positive associations between intake of sugar-sweetened beverages, ultra-processed foods and higher risk of CD but not UC. Although genetic studies have reported

overlapped genetic risk alleles, disease—specific alleles still exist for UC and CD, which might mediate disease-specific gene-environmental effects.⁴⁰ Overall, the mechanisms underlying the disparity of the effect of air pollutants on UC and CD were unclear enough and merited further investigation.

Specific air pollutants such as PM, O_3 , NO_2 , and polycyclic aromatic hydrocarbons have been reported to influence DNA methylation of genes related to multiple chronic diseases.²⁵ By utilizing genetic instrumental variables of DNA methylation markers associated with NO_2 as well as $PM_{2.5}$ and building upon the largest IBD GWAS, our MR analyses identified significant associations between modified methylation at multiple air pollution-related CpG sites and the susceptibility of UC. The roles of *CXCR2* and MHC class III region loci were further validated when we compared the findings with genome-wide methylation analysis of UC and conducted

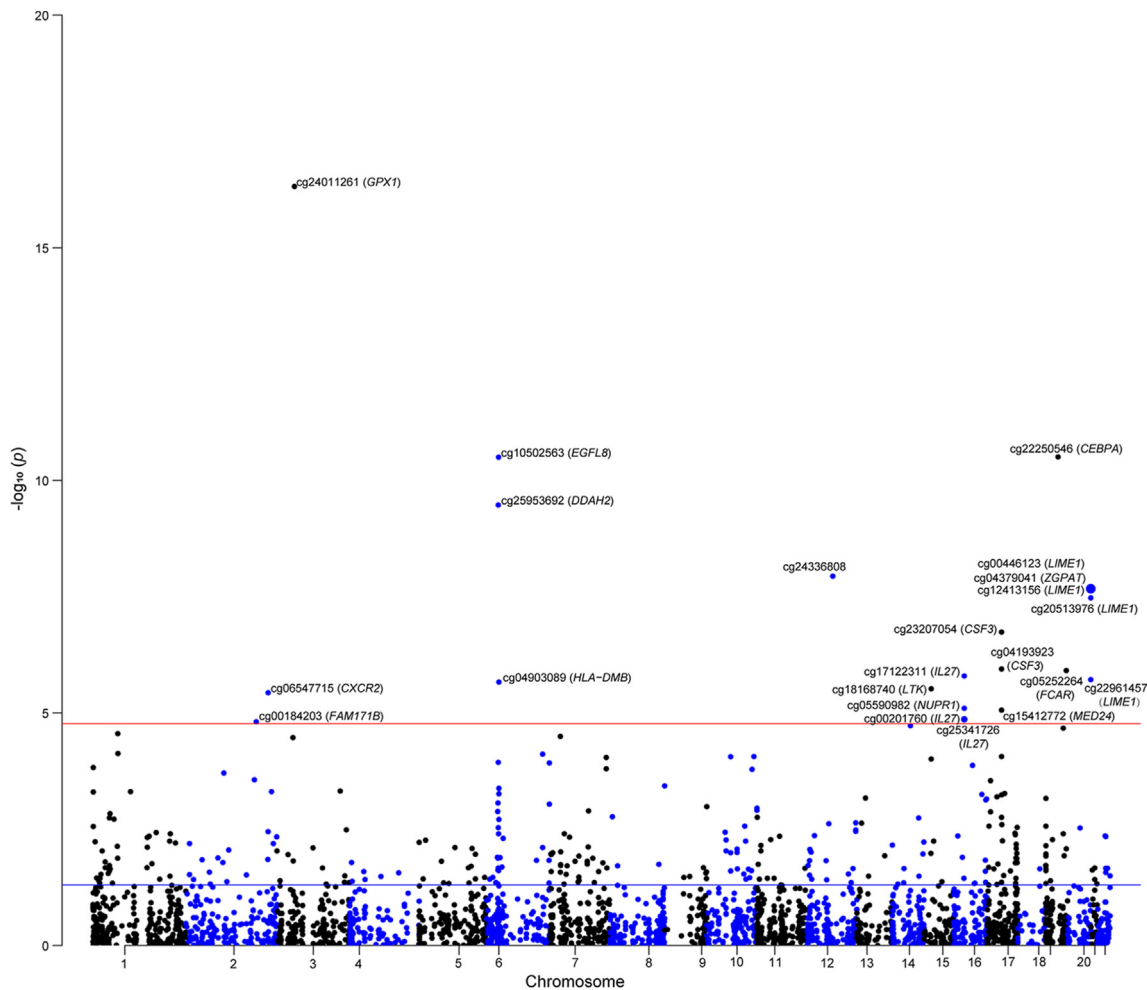


Fig. 3: The effect of gene regulated by NO₂-related DNA methylation on the risk of ulcerative colitis. The blue line means P = 0.05, and the red line means the threshold of FDR correction.

colocalization analysis. Importantly, we also examined the associations between DNA methylation and gene expression in colonic tissue and performed a single cell RNA sequencing analysis to investigate the expression of these two genes in distinct cell types of the intestinal tissues.

CXC motif chemokine receptor 2 (CXCR2) is the common receptor of CXCL1, CXCL5 and CXCL8, which are pro-inflammatory cytokines expressed in IBD.

Consistently, a genome-wide DNA methylation study conducted using colon mucosal biopsies from treatment-naïve patients with UC and controls has confirmed our finding regarding CXCR2. The study revealed differential expression of CXCR2 in the mucosa of patients with UC and its association with observed differential DNA methylation (P = 0.016).⁵⁷ CXCR2 has been demonstrated to be essential for the migration of neutrophils after air pollutant exposure.⁵⁸ As for the role of CXCR2 in UC, data based on samples from populations⁵⁹ and animal models⁶⁰ have reported increased expression of CXCR2 in active colonic IBD and support the hypothesis of the important pathogenic role of CXCR2+ neutrophils in the development of colitis.

Besides CXCR2, we discovered the roles of MHC class III region on associations between air pollution and incident UC. The MHC region has been extensively researched in the human genome due to its significant associations with autoimmune and inflammatory

Exposure	CpG site	Gene	Beta	SE	P-value
PM _{2.5}	cg16689962	AGPAT1	2.279	0.212	5.84E-27
NO ₂	cg06547715	CXCR2	0.096	0.021	3.68E-06

UC, ulcerative colitis; NO₂, nitrogen dioxide; PM, particulate matter; SE, standard.

Table 4: The loci validated in the genome-wide methylation analysis of ulcerative colitis.

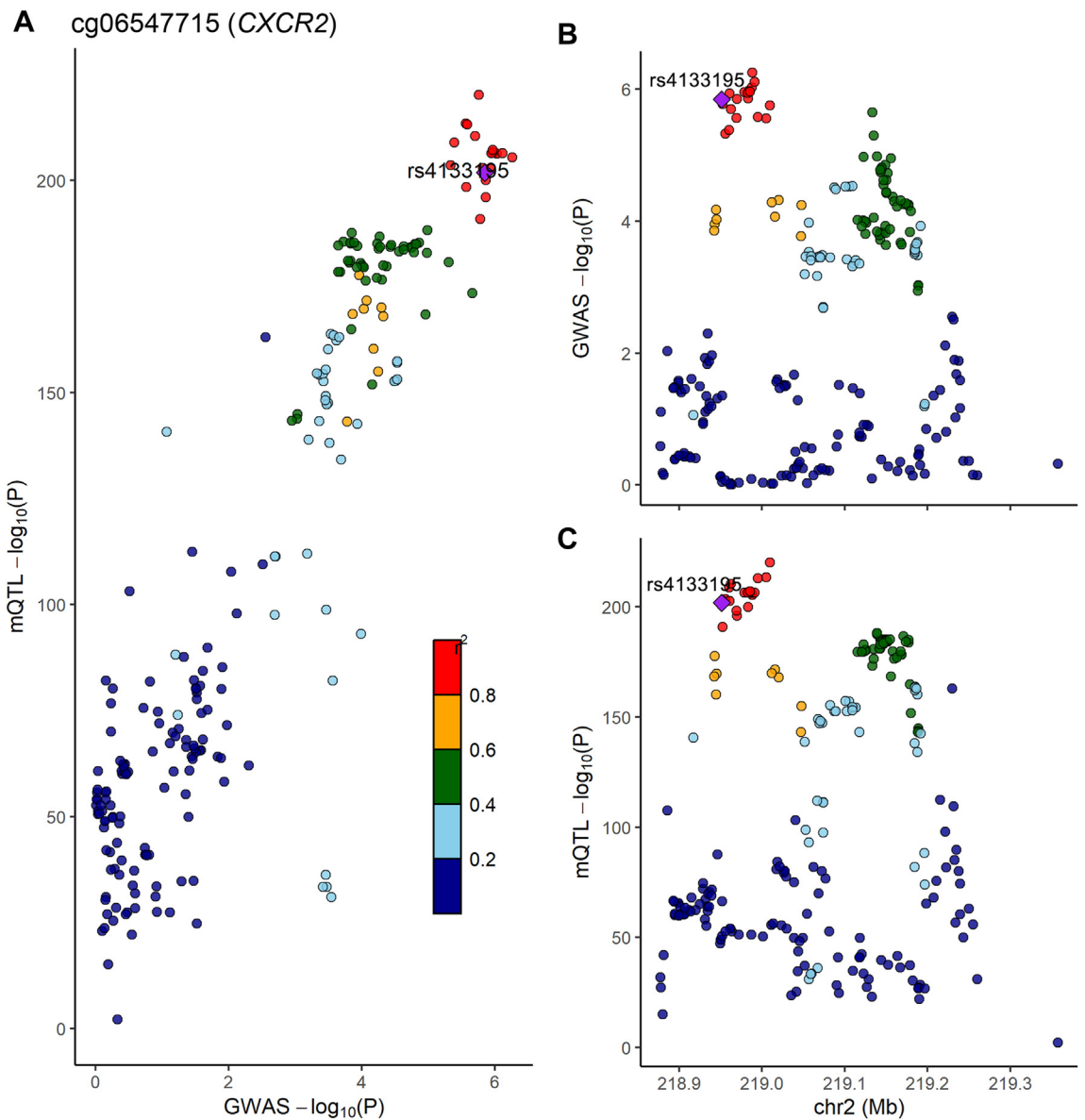


Fig. 4: Regional plot of colocalization evidence of CpG site methylation and ulcerative colitis susceptibility. Panel A displays the P -value of the SNP and mQTL in corresponding GWAS and EWAS. In the Panel B and C, the horizontal axis demonstrates the base position of the SNP and mQTL, and the vertical axis is the P -value of the SNP and mQTL, respectively.

diseases, making it one of the most studied region in this field.⁶¹ Interestingly, early linkage and association studies had indicated that the region might have a more significant impact on susceptibility to UC than CD,⁶² with specific allelic associations being reported and validated for disease severity and extent. Goyette et al. conducted the most comprehensive analysis to date, which involved SNP typing of the MHC in over 32,000 individuals with IBD. Their findings demonstrated the presence of multiple human leukocyte antigen (HLA) alleles, with a predominant role observed for HLA-

DRB1*01:03 in both UC and CD.⁶³ A recent study has also identified a distinct DNA methylation pattern in the MHC region of the sigmoid colon in patients with IBD.⁶⁴ Further studies are needed to explore the interplay of genetic and epigenetic variation across the MHC III region, and to elucidate the mechanisms involved in the associations of these variations with the exposome in the pathogenesis of IBD and other autoimmune diseases. The close clustering of genes and linkage disequilibrium across the region are important considerations in study design and interpretation.

In addition to the *CXCR2* and MHC class III region loci, a series of genes including *SMAD3*, *P4HA2*, *NCR3*, *GPX1*, *EGFL8*, *CEBPA*, *DDAH2*, *LIME1* were discovered in the first stage of our study; it is possible that these loci might also be involved in linking air pollution and the risk of UC, although they were not validated in the additional analyses. Gene sites such as *SMAD3*⁶⁵ have been identified by previous studies as IBD susceptibility loci, which might strengthen their potential role in the disease. Of interest, *GPX-1* is an intracellular antioxidant enzyme, limiting the harmful effects of hydrogen peroxide,⁶⁶ thus emphasizing its role in modulating cellular oxidant stress. In the mouse, combined deficiency of *GPX1* and *GPX2* pre-disposes to ileocolitis.^{67,68} Notably, the NO₂-related methylated CpG sites cg07462448 and cg24084564 have been validated in an epigenome-wide analysis of the association between air pollution and DNA methylation, which showed positive associations between methylation status at these determinants and NO_x as well as light-absorbing carbon in the multi-ethnic US born-participants.⁶⁹

Importantly, smoking has been found to be associated with alterations of methylation at *CXCR2*, *SMAD3* and *GPX1*⁷⁰⁻⁷² and loci at MHC class III region.^{9,73} *CXCR2* has been demonstrated to play an essential role in smoke-induced inflammation, and inhibition of *CXCR2* can reduce neutrophilic infiltration and tissue damage.⁷¹ In addition, cigarette smoke leads to the increase of *WNT-5B* expression, thus increasing *TGF-β/SMAD3*⁷² signalling in airway epithelial cells. Interestingly, epigenetic changes in *LTA/TNF* loci, located within the MHC III class region, were also found to play an important role in the association between smoking and CD risk.⁹ This suggests that the DNA methylation within MHC III class region may be influenced by multiple factors and involved in both pathogenesis of UC and CD. Pathways involving secretory vesicles, lysosome, leukocyte activation, and a series of immunity responses were identified in our gene ontology analysis. Overall, investigation of ontology of CpG sites related to NO₂ exposure, along with analyses of transcriptomic correlates of MR-identified CpG sites, highlighted the broad impact of air pollution on immunity. Furthermore, an RCT study offers valuable insights for guiding the future direction of intervention research in patients with IBD, building upon the current findings.³²

Our study has several strengths. This is a comprehensive study to explore the potential causal associations between air pollutants and the risk of IBD, and potential mechanisms from the perspective of DNA methylation. Its design triangulated the evidence across observational and genetic analyses, and across samples from colon tissue providing robust evidence for the mediating effects of DNA methylation on air pollution and UC. Of note, there are certain limitations in our current study. First, given the specific cohort under investigation, our study focused exclusively on participants of European

ancestry and older age, thus limiting the generalizability of our findings to other populations. Further research is required to confirm these findings, necessitating the inclusion of diverse populations due to variations in methylation patterns among ethnic sub-groups.⁷⁴ Secondly, although the LUR model has been recognized as an effective approach for estimating air pollution exposure, particularly for NO₂ and NO_x, there may still be potential misclassification. This could be attributed to the fact that the simplistic dispersion assumption of the LUR model is more suitable for traffic emissions rather than industrial point emissions. Additionally, factors such as the number of monitoring sites and the complexity and size of urban environments can also influence its predictive accuracy.³⁵ Nevertheless, these challenges can be addressed through the implementation of an MR design which is less susceptible to environmental influences.

Although our MR analysis compensated for the bias of observational studies, pleiotropy may exist when using single nucleotide polymorphisms as instrumental variables. In the analyses for testing horizontal pleiotropy, we found no horizontal pleiotropy for PM_{2.5} but we detected horizontal pleiotropy for the analysis of NO₂. It is important to note that we studied one instrument variable for most of air pollution related CpG sites, and we have not been able to assess these results by performing multivariable MR and other sensitivity analyses. However, we have supplemented our data with a series of validation analyses to confirm the findings.

In addition, we acknowledge that the evaluation of air pollution and the associated changes in DNA methylation CpG sites relied on cross-sectional data, thereby lacking longitudinal monitoring of exposure and the dynamic changes of exposure on methylation. Finally, although there is evidence suggesting that certain methylation markers in the blood may reflect corresponding methylation signatures in intestinal tissues,⁷⁵ it should be noted that DNA methylation characteristics vary among different cell types and tissues.⁷⁶ That means that the current study has not directly investigated the altered DNA methylation patterns in intestinal tissues induced by air pollution. Again, we have substantiated the findings related to these two target genes through a series of additional analyses. These includes an investigation into the regulatory pattern of DNA methylation on gene expression in colonic tissues, querying the GTEx Portal, and conducting single cell RNA sequencing analysis on distinct intestinal cells. Furthermore, the key findings have biological plausibility at several levels; the discovery regarding *CXCR2* has also been corroborated by a previous genome-wide DNA methylation study conducted on mucosal samples⁵⁷ Biomarkers identified in both blood and intestinal tissue serve as complementary tools for identifying diagnostic and treatment targets for the disease. The above issue should be more directly

addressed in future studies, particularly with the availability of specific EWAS of air pollution conducted across various tissues and cell types.

In summary, our study provides an exciting research paradigm for future studies that aim to explore potential mechanisms of environment-IBD by integrating observational, genetic and methylation evidence and highlights the importance of epigenetic alterations in complex disease pathogenesis of UC. We report a potential causal association between air pollution and risk of UC, but not CD. We also provide evidence for biological mechanisms for this association, specifically the epigenetic alteration of *CXCR2* and determinants within the MHC class III region. The findings of epigenetic dysregulation linking air pollution to the risk of UC demonstrate the translational potential of these techniques in defining biomarkers as well as characterizing pathways for pharmacological interventions.

Contributors

Jie Chen, methodology (lead), formal analysis (equal), writing-original draft (equal); Han Zhang, methodology (equal), formal analysis (lead), writing-original draft (equal); Tian Fu, methodology (equal), formal analysis (equal), writing-original draft (lead); Jianhui Zhao, methodology (supporting), writing-original draft (supporting), writing-review and editing (supporting); Jan Krzysztof Nowak, formal analysis (equal), conceptualization (equal), supervision (equal); Rahul Kalla, methodology (equal), writing-review and editing (equal); Judith Wellens, methodology (equal), writing-review and editing (lead); Shuai Yuan, methodology (supporting), writing-review and editing (supporting); Alexandra Noble, methodology (equal), writing-review and editing (supporting); Nicholas T Ventham, conceptualization (equal), writing-review and editing (supporting); Malcolm Dunlop, writing-review and editing (supporting), methodology (equal); Jonas Halfvarson, conceptualization (supporting), writing-review and editing (supporting); Ren Mao, methodology (equal), writing-review and editing (equal), supervision (equal); Evropi Theodoratou, conceptualization (lead), supervision (supporting), writing-review and editing (equal); Jack Satsangi, conceptualization (lead), supervision (equal), writing-review and editing (equal); Xue Li, conceptualization (lead), methodology (lead), supervision (lead), writing-review and editing (supporting). Jie Chen, Han Zhang, Tian Fu, and Xue Li have directly accessed and verified the underlying data reported in the manuscript. Evropi Theodoratou, Jack Satsangi, and Xue Li are the study guarantors. All authors read and gave final approval of the version to be published.

Data sharing statement

The results of this study are included in this article and the Supporting files. The UK Biobank is an open access resource and researchers required approval from the UK Biobank (www.ukbiobank.ac.uk/). Genetic data was derived from the large genome-wide association study (GWAS), Genetics of DNA Methylation Consortium (GoDMC) and genome-wide DNA methylation analyses, which can be required from the published articles.

Declaration of interests

JKN reports consulting for Procter & Gamble and grant support from the Biocodex Microbiota Foundation, outside of the submitted work. All the remaining authors declare no potential or actual conflict of interest to the work presented in this paper.

Acknowledgements

We appreciate the participants and managing team of UK Biobank [Application number of this study: 73595] and all investigators for sharing these data. Our work was supported by the Natural Science

Fund for Distinguished Young Scholars of Zhejiang Province (LR22H260001) [XL] and the National Nature Science Foundation of China (No. 82204019) [XL]; CRUK Career Development Fellowship (C31250/A22804) [ET]; and by the Research Foundation Flanders (FWO), Belgium by a PhD Fellowship strategic basic research (SB) grant (1S06023N) [JW]; National Science Center, Poland (No. 2020/39/D/NZ5/02720) [JKN]. The IBD Character was supported by the European Union's Seventh Framework Programme [FP7] grant IBD Character (No. 2858546).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2024.105443>.

References

- Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet*. 2017;390(10114):2769–2778.
- Kuenzig ME, Fung SG, Marderfeld L, et al. Twenty-first century trends in the global epidemiology of pediatric-onset inflammatory bowel disease: systematic review. *Gastroenterology*. 2022;162(4):1147–1159.e4.
- Jones GR, Lyons M, Plevris N, et al. IBD prevalence in Lothian, Scotland, derived by capture-recapture methodology. *Gut*. 2019;68(11):1953–1960.
- Liu Z, Liu R, Gao H, et al. Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat Genet*. 2023;55(5):796–806.
- Ordás I, Eckmann L, Talamini M, Baumgart DC, Sandborn WJ. Ulcerative colitis. *Lancet*. 2012;380(9853):1606–1619.
- Sun Y, Yuan S, Chen X, et al. The contribution of genetic risk and lifestyle factors in the development of adult-onset inflammatory bowel disease: a prospective cohort study. *Am J Gastroenterol*. 2023;118(3):511–522.
- Chen J, Wellens J, Kalla R, et al. Intake of ultra-processed foods is associated with an increased risk of crohn's disease: a cross-sectional and prospective analysis of 187 154 participants in the UK biobank. *J Crohns Colitis*. 2023;17(4):535–552.
- Fu T, Chen H, Chen X, et al. Sugar-sweetened beverages, artificially sweetened beverages and natural juices and risk of inflammatory bowel disease: a cohort study of 121,490 participants. *Aliment Pharmacol Ther*. 2022;56(6):1018–1029.
- Zhang H, Kalla R, Chen J, et al. Altered DNA methylation within DNMT3A, AHRH, LTA/TNF loci mediates the effect of smoking on inflammatory bowel disease. *Nat Commun*. 2024;15(1):595.
- Noble AJ, Nowak JK, Adams AT, Uhlig HH, Satsangi J. Defining interactions between the genome, epigenome, and the environment in inflammatory bowel disease: progress and prospects. *Gastroenterology*. 2023;165(1):44–60.e2.
- Ananthakrishnan AN, Bernstein CN, Iliopoulos D, et al. Environmental triggers in IBD: a review of progress and evidence. *Nat Rev Gastroenterol Hepatol*. 2018;15(1):39–49.
- Elten M, Benchimol EI, Fell DB, et al. Ambient air pollution and the risk of pediatric-onset inflammatory bowel disease: a population-based cohort study. *Environ Int*. 2020;138:105676.
- Adami G, Pontalti M, Cattani G, et al. Association between long-term exposure to air pollution and immune-mediated diseases: a population-based cohort study. *RMD Open*. 2022;8(1):e002055.
- Opstelten JL, Beelen RMJ, Leenders M, et al. Exposure to ambient air pollution and the risk of inflammatory bowel disease: a European nested case-control study. *Dig Dis Sci*. 2016;61(10):2963–2971.
- Kaplan GG, Hubbard J, Korzenik J, et al. The inflammatory bowel diseases and ambient air pollution: a novel association. *Am J Gastroenterol*. 2010;105(11):2412–2419.
- Li FR, Wu KY, Fan WD, Chen GC, Tian H, Wu XB. Long-term exposure to air pollution and risk of incident inflammatory bowel disease among middle and old aged adults. *Ecotoxicol Environ Saf*. 2022;242:113835.
- Huang Y, Zhu M, Ji M, et al. Air pollution, genetic factors, and the risk of lung cancer: a prospective study in the UK biobank. *Am J Respir Crit Care Med*. 2021;204(7):817–825. <https://doi.org/10.1164/rccm.202011-4063OC>.
- Wang M, Zhou T, Song Q, et al. Ambient air pollution, healthy diet and vegetable intakes, and mortality: a prospective UK Biobank study. *Int J Epidemiol*. 2022;51(4):1243–1253. <https://doi.org/10.1093/ije/dyac022>.

- 19 Fu Z, Liu Q, Liang J, et al. Air pollution, genetic factors and the risk of depression. *Sci Total Environ.* 2022;850:158001. <https://doi.org/10.1016/j.scitotenv.2022.158001>.
- 20 Gao X, Jiang M, Huang N, Guo X, Huang T. Long-Term air pollution, genetic susceptibility, and the risk of depression and anxiety: a prospective study in the UK biobank cohort. *Environ Health Perspect.* 2023;131(1):17002. <https://doi.org/10.1289/EHP0391>.
- 21 Li D, Ma Y, Cui F, et al. Long-term exposure to ambient air pollution, genetic susceptibility, and the incidence of bipolar disorder: a prospective cohort study. *Psychiatry Res.* 2023;327:115396. <https://doi.org/10.1016/j.psychres.2023.115396>.
- 22 Zhang J, Fang XY, Wu J, et al. Association of combined exposure to ambient air pollutants, genetic risk, and incident rheumatoid arthritis: a prospective cohort study in the UK biobank. *Environ Health Perspect.* 2023;131(3):37008. <https://doi.org/10.1289/EHP10710>.
- 23 Ma Y, Li D, Cui F, et al. Air pollutants, genetic susceptibility, and abdominal aortic aneurysm risk: a prospective study. *Eur Heart J.* 2024;45(12):1030–1039. <https://doi.org/10.1093/eurheartj/ehad886>.
- 24 Ventham NT, Kennedy NA, Adams AT, et al. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat Commun.* 2016;7:13507.
- 25 Poursafa P, Kamali Z, Fraszczyk E, Boecker HM, Vaez A, Snieder H. DNA methylation: a potential mediator between air pollution and metabolic syndrome. *Clin Epigenetics.* 2022;14(1):82.
- 26 Rider CF, Carlsen C. Air pollution and DNA methylation: effects of exposure in humans. *Clin Epigenetics.* 2019;11(1):131.
- 27 Kalla R, Adams AT, Nowak JK, et al. Analysis of systemic epigenetic alterations in inflammatory bowel disease: defining geographical, genetic and immune-inflammatory influences on the circulating methylome. *J Crohns Colitis.* 2023;17(2):170–184.
- 28 Ventham NT, Kennedy NA, Nimmo ER, Satsangi J. Beyond gene discovery in inflammatory bowel disease: the emerging role of epigenetics. *Gastroenterology.* 2013;145(2):293–308.
- 29 Ventham NT, Kennedy NA, Kalla R, et al. Genome-wide methylation profiling in 229 patients with Crohn's disease requiring intestinal resection: epigenetic analysis of the trial of prevention of post-operative crohn's disease (TOPPIC). *Cell Mol Gastroenterol Hepatol.* 2023;16(3):431–450.
- 30 Joustra V, Li Yim A, YF, Hageman I, et al. Long-term temporal stability of peripheral blood DNA methylation profiles in patients with inflammatory bowel disease. *Cell Mol Gastroenterol Hepatol.* 2023;15(4):869–885.
- 31 Joustra V, Hageman IL, Satsangi J, et al. Systematic review and meta-analysis of peripheral blood DNA methylation studies in inflammatory bowel disease. *J Crohns Colitis.* 2023;17(2):185–198.
- 32 Sun Y, Huang J, Zhao Y, et al. Inflammatory cytokines and DNA methylation in healthy young adults exposure to fine particulate matter: a randomized, double-blind crossover trial of air filtration. *J Hazard Mater.* 2020;398:122817. <https://doi.org/10.1016/j.jhazmat.2020.122817>.
- 33 Sekula P, Del Greco MF, Pattaro C, Köttgen A. Mendelian randomization as an approach to assess causality using observational data. *J Am Soc Nephrol.* 2016;27(11):3253–3265.
- 34 Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779. <https://doi.org/10.1371/journal.pmed.1001779>.
- 35 de Hoogh K, Korek M, Vienneau D, et al. Comparing land use regression and dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies. *Environ Int.* 2014;73:382–392. <https://doi.org/10.1016/j.envint.2014.08.011>.
- 36 Eeftens M, Beelen R, De Hoogh K, et al. Development of land use regression models for PM_{2.5}, PM_{2.5} absorbance, PM₁₀ and PM_{coarse} in 20 European study areas; results of the ESCAPE project. *Environ Sci Technol.* 2012;46(20):11195–11205.
- 37 Beelen R, Hoek G, Vienneau D, et al. Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe—The ESCAPE project. *Atmos Environ.* 2013;72:10–23.
- 38 Li J, Lu A, Si S, et al. Exposure to various ambient air pollutants increases the risk of venous thromboembolism: a cohort study in UK Biobank. *Sci Total Environ.* 2022;845:157165. <https://doi.org/10.1016/j.scitotenv.2022.157165>.
- 39 Fu T, Ye S, Sun Y, Dan L, Wang X, Chen J. Greater adherence to cardioprotective diet can reduce inflammatory bowel disease risk: a longitudinal cohort study. *Nutrients.* 2022;14(19):4058. <https://doi.org/10.3390/nu14194058>.
- 40 Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet.* 2015;47(9):979–986.
- 41 Andersson T, Alfredsson L, Källberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. *Eur J Epidemiol.* 2005;20(7):575–579.
- 42 Panni T, Mehta AJ, Schwartz JD, et al. Genome-wide analysis of DNA methylation and fine particulate matter air pollution in three study populations: KORA F3, KORA F4, and the normative aging study. *Environ Health Perspect.* 2016;124(7):983–990. <https://doi.org/10.1289/ehp.1509966>.
- 43 de F C Lichtenfels AJ, van der Plaats DA, de Jong K, et al. Long-term air pollution exposure, genome-wide DNA methylation and lung function in the LifeLines cohort study. *Environ Health Perspect.* 2018;126(2):027004. <https://doi.org/10.1289/EHP2045>.
- 44 Min JL, Hemani G, Hannon E, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet.* 2021;53(9):1311–1321. <https://doi.org/10.1038/s41588-021-00923-x>.
- 45 Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet.* 2000;25(1):25–29.
- 46 Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics.* 2016;32(2):286–288.
- 47 de Lange KM, Moutsianas L, Lee JC, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet.* 2017;49(2):256–261. <https://doi.org/10.1038/ng.3760>.
- 48 Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol.* 2015;44(2):512–525. <https://doi.org/10.1093/ije/dyv080>.
- 49 Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10(5):e1004383. <https://doi.org/10.1371/journal.pgen.1004383>.
- 50 Xiong Z, Li M, Yang F, et al. EWAS Data Hub: a resource of DNA methylation array data and metadata. *Nucleic Acids Res.* 2020;48(D1):D890–D895. <https://doi.org/10.1093/nar/gkz840>.
- 51 Xiong Z, Yang F, Li M, et al. EWAS Open Platform: integrated data, knowledge and toolkit for epigenome-wide association study. *Nucleic Acids Res.* 2022;50(D1):D1004–D1009. <https://doi.org/10.1093/nar/gkab972>.
- 52 Michaux M, Chan JM, Bergmann L, Chaves LF, Klinkenberg B, Jacobson K. Spatial cluster mapping and environmental modeling in pediatric inflammatory bowel disease. *World J Gastroenterol.* 2023;29(23):3688–3702.
- 53 Fecht D, Fischer P, Fortunato L, et al. Associations between air pollution and socioeconomic characteristics, ethnicity and age profile of neighbourhoods in England and The Netherlands. *Environ Pollut.* 2015;198:201–210.
- 54 Armitage EL, Aldhous MC, Anderson N, et al. Incidence of juvenile-onset Crohn's disease in Scotland: association with northern latitude and affluence. *Gastroenterology.* 2004;127(4):1051–1057.
- 55 Piovani D, Danese S, Peyrin-Biroulet L, Nikolopoulos GK, Lytras T, Bonovas S. Environmental risk factors for inflammatory bowel diseases: an umbrella review of meta-analyses. *Gastroenterology.* 2019;157(3):647–659.e4.
- 56 Yuan S, Sun Y, Tan X, et al. Sleep duration and daytime napping in relation to incident inflammatory bowel disease: a prospective cohort study. *Aliment Pharmacol Ther.* 2023;57(5):475–485.
- 57 Taman H, Fenton CG, Hensel IV, Anderssen E, Florholmen J, Paulssen RH. Genome-wide DNA methylation in treatment-naïve ulcerative colitis. *J Crohns Colitis.* 2018;12(11):1338–1347. <https://doi.org/10.1093/ecco-jcc/jjy117>.
- 58 Ha H, Debnath B, Neamati N. Role of the CXCL8-CXCR1/2 Axis in cancer and inflammatory diseases. *Theranostics.* 2017;7(6):1543–1588.
- 59 Kvedaraitė E, Lourda M, Ideström M, et al. Tissue-infiltrating neutrophils represent the main source of IL-23 in the colon of patients with IBD. *Gut.* 2016;65(10):1632–1641.
- 60 Ajuebor MN, Zagorski J, Kunkel SL, Strieter RM, Hogaboam CM. Contrasting roles for CXCR2 during experimental colitis. *Exp Mol Pathol.* 2004;76(1):1–8.

- 61 Fernando MM, Stevens CR, Walsh EC, et al. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet.* 2008;4(4):e1000024.
- 62 Satsangi J, Welsh KI, Bunce M, et al. Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. *Lancet.* 1996;347(9010):1212–1217.
- 63 Goyette P, Boucher G, Mallon D, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet.* 2015;47(2):172–179.
- 64 Dennison TW, Edgar RD, Payne F, et al. Patient-derived organoid biobank identifies epigenetic dysregulation of intestinal epithelial MHC-I as a novel mechanism in severe Crohn's disease. *Gut.* 2024;10(332043):2024–332043.
- 65 Lees CW, Barrett JC, Parkes M, Satsangi J. New IBD genetics: common pathways with other diseases. *Gut.* 2011;60(12):1739–1753.
- 66 Lubos E, Loscalzo J, Handy DE. Glutathione peroxidase-1 in health and disease: from molecular mechanisms to therapeutic opportunities. *Antioxid Redox Signal.* 2011;15(7):1957–1997.
- 67 Esworthy RS, Kim BW, Larson GP, et al. Colitis locus on chromosome 2 impacting the severity of early-onset disease in mice deficient in GPX1 and GPX2. *Inflamm Bowel Dis.* 2011;17(6):1373–1386.
- 68 Chu FF, Esworthy RS, Doroshov JH, et al. Deficiency in Duox2 activity alleviates ileitis in GPx1- and GPx2-knockout mice without affecting apoptosis incidence in the crypt epithelium. *Redox Biol.* 2017;11:144–156.
- 69 Watkins SH, Testa C, Simpkin AJ, et al. An epigenome-wide analysis of DNA methylation, racialized and economic inequities, and air pollution. *bioRxiv.* 2023. <https://doi.org/10.1101/2023.12.07.570610>.
- 70 Beane J, Cheng L, Soldi R, et al. SIRT1 pathway dysregulation in the smoke-exposed airway epithelium and lung tumor tissue. *Cancer Res.* 2012;72(22):5702–5711.
- 71 Thatcher TH, McHugh NA, Egan RW, et al. Role of CXCR2 in cigarette smoke-induced lung inflammation. *Am J Physiol Lung Cell Mol Physiol.* 2005;289(2):L322–L328.
- 72 Heijink IH, de Bruin HG, Dennebos R, et al. Cigarette smoke-induced epithelial expression of WNT-5B: implications for COPD. *Eur Respir J.* 2016;48(2):504–515.
- 73 Füst G, Arason GJ, Kramer J, et al. Genetic basis of tobacco smoking: strong association of a specific major histocompatibility complex haplotype on chromosome 6 with smoking behavior. *Int Immunol.* 2004;16(10):1507–1514.
- 74 Galanter JM, Gignoux CR, Oh SS, et al. Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. *Elife.* 2017;6:e20532.
- 75 Zhang Y, Bewerunge-Hudler M, Schick M, et al. Blood-derived DNA methylation predictors of mortality discriminate tumor and healthy tissue in multiple organs. *Mol Oncol.* 2020;14(9):2111–2123. <https://doi.org/10.1002/878-0261.12738>.
- 76 Wang X, Campbell MR, Cho HY, Pittman GS, Martos SN, Bell DA. Epigenomic profiling of isolated blood cell types reveals highly specific B cell smoking signatures and links to disease risk. *Clin Epigenetics.* 2023;15(1):90. <https://doi.org/10.1186/s13148-023-01507-8>.