CROHN'S
& COLITIS
FOUNDATION

An Atlas Characterizing the Shared Genetic Architecture of Inflammatory Bowel Disease with Clinical and Behavioral Traits

Vikram R. Shaw, MS,[*] Jinyoung Byun, PhD,[*,†,‡] Rowland W. Pettit, MD, PhD,[*] Jason K. Hou, MD, MS,[§] Kyle M. Walsh, PhD,[‖] Younghun Han, PhD,[*,†] and Christopher I. Amos, PhD[*,†,‡]

From the [*]Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX, USA
[†]Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX, USA
[‡]Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA
[§]Department of Medicine-Gastroenterology, Baylor College of Medicine, Houston, TX, USA
[‖]Division of Neuro-epidemiology, Department of Neurosurgery, Duke University School of Medicine, Durham, NC, USA
Address correspondence to: Christopher I. Amos, Institute for Clinical and Translational Research, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas, 77030, USA. Tel: 713-798-3394; Fax: 713-798-1499 (Chris.Amos@bcm.edu).

Background: Inflammatory bowel disease (IBD) development is a complex, multifactorial process that involves extrinsic and intrinsic factors such as host genetics, the immune system, the gut microbiome, and environmental risks. To help understand the genetic contribution of clinical, behavioral, psychiatric, and diet-related traits, we aim to provide a deep and comprehensive characterization of the shared genetic architecture between IBD and hundreds of potentially related traits.

Methods: Utilizing publicly available summary statistics from a previously published IBD genome-wide association study and hundreds of traits from the United Kingdom BioBank (UKBB), we performed linkage disequilibrium score regression (LDSR) analysis to estimate cross-trait genetic correlations between Crohn's disease (CD), ulcerative colitis (UC), and IBD summary statistics with the UKBB traits of interest.

Results: Nominally significant ($P < .05$) genetic correlations were observed for 181 traits in overall IBD, 239 traits in CD, and 94 traits in UC. We replicate the known association between smoking behavior and CD/UC, namely that current tobacco smoking has a positive genetic correlation with CD (rg = 0.12, $P = 4.2 \times 10^{-4}$), while "ever smoking" has a negative genetic correlation with UC (rg = −0.07, $P = .042$). Globally, all 3 strata (IBD, CD, and UC) demonstrated increased genetic correlations for psychiatric-related traits related to anxiety and depression.

Conclusion: The present analysis reveals the shared genetic architecture between multiple traits and IBD, CD, and UC. Understanding the relevance of joint occurrences of IBD with psychiatric diseases may moderate management of these diseases for individuals jointly affected by them.

Lay Summary

This study provides an atlas of the genetic correlation between hundreds of United Kingdom Biobank (UKBB) traits with inflammatory bowel disease (IBD), Crohn's disease (CD), and ulcerative colitis (UC). Notable strong correlations are seen between IBD and various psychiatric traits.

Key Words: genetic correlation, GWAS, psychiatric comorbidities

Introduction

Inflammatory bowel disease (IBD) is a worldwide healthcare problem that is characterized by chronic, relapsing intestinal inflammation. In North America, the estimated prevalence of Crohn's disease (CD) and ulcerative colitis (UC), the 2 main forms of IBD, are around 300 per 100 000 people.[1] Inflammatory bowel disease is often diagnosed in early adulthood and can lead to a decreased quality of life, in addition to increased risks of anxiety and depression.[2] Additionally, the life expectancy of patients with IBD is reduced by 5.0 to 8.1 years compared with individuals without IBD.[3] Despite treatment advances in recent years, many patients with IBD experience severe disease that requires aggressive or invasive treatment, such as total parental nutrition (TPN) or extensive surgery, highlighting the

variable response of particular groups of patients to first- and second-line treatments.[4]

Part of this variability may be explained by the fact that IBD is a heterogeneous pathology along several axes.[5] Patients with CD, for example, may experience diverse clinical manifestations with regard to disease location and behavior, disease progression, perianal involvement, extraintestinal manifestations, histopathology, and disease response to various therapies.[6] This heterogeneity may be explained by the complex development of IBD, which is a multifactorial process that involves extrinsic and intrinsic factors, such as host genetics, the immune system, the gut microbiome, and environmental risks.[7,8] The role of genetics in IBD development is well-established, with up to 12% of IBD patients having a family history of IBD.[9] Family and twin studies have

---

Key Messages

- Epidemiological studies have individually characterized the roles of risk factor groups (eg, smoking behavior, diet, etc.) in the development and progression of inflammatory bowel disease.
- Our study provides a comprehensive atlas of genetically correlated traits of different categories (eg, diet-related, psychiatric, comorbid clinical conditions) with inflammatory bowel disease, Crohn's disease, and ulcerative colitis.
- A significant positive genetic correlation is observed between inflammatory bowel disease and psychiatric disease burden.

---

shown that host genetics play a greater role in CD compared with UC etiology, with first-degree relatives of CD patients experiencing greater risk of IBD than first-degree relatives of UC patients.[9] Since the first genome-wide association study (GWAS) of IBD was performed in 2005,[10] additional GWAS have identified >230 single nucleotide polymorphisms (SNPs) associated with IBD and helped to characterize the roles of the immune system in IBD development.[11-13] For example, altered epithelial barrier function contributes to intestinal inflammation in patients with UC, and altered innate immune responses are more closely associated with CD development.[14]

Although genetic susceptibility is an important IBD risk factor, environmental risk factors, such as diet or smoking behaviors, also play an important role.[15] Associated environmental factors encompass a breadth of potential risk factors, ranging from early-life exposures (eg, mode of childbirth, breastfeeding, antibiotic exposure) to later-life exposures (eg, smoking, life stressors, diet, lifestyle).[15] Data point to a differential association between smoking and CD vs UC, with current smoking positively associated with CD and smoking cessation positively associated with UC.[15] Diet is another important factor in IBD risk, with saturated fats and low vitamin D levels associated with increased incident IBD.[16]

The current IBD treatment landscape includes a broad therapeutic armamentarium, but the current treatment algorithm may not reflect all of the advances made in understanding the genetic and environmental etiology of IBD.[17] Providing personalized care to patients will depend on parsing through the heterogeneity of the diseases to identify predictors for disease course and therapeutic response.[17] To help decompose this heterogeneity, we aim to provide a deeper and more comprehensive characterization of the shared genetic architecture between IBD overall, and CD and UC individually, with hundreds of potentially related traits using GWAS summary statistics. Focusing on understanding the genetic architecture behind IBD codevelopment with other traits may provide additional information compared with studying individual phenotypes in isolation.[18] While other studies have performed epidemiological and genetic investigations into particular traits of interest, to our knowledge no other study has performed wide-scale, cross-trait analysis of hundreds of clinical, behavioral, psychiatric, and diet-related traits with IBD risk. The present study uses a statistical framework known as cross-trait linkage disequilibrium score regression (LDSR) to fill this gap in knowledge. Linkage disequilibrium score regression utilizes GWAS summary statistic data to identify genome-wide genetic correlations across traits of interest, using SNP

effect estimates to calculate genetic cocorrelation (r_g) between traits.[19] As has previously been shown in lung cancer, using this method to investigate the disease of interest (IBD) in the context of hundreds of potential disease codevelopment traits can provide important etiological insights into the disease of interest.[18] In the present study, we quantify the association between genetically influenced clinical, behavioral, psychiatric, and diet-related traits and the risk of IBD, CD, or UC using publicly available summary statistics from prior IBD GWAS and LDSRs to estimate cross-trait genetic correlations. We aim to confirm prior associations between important risk factor or outcome traits with IBD and to identify potentially novel phenotype-trait associations.

Materials and Methods

Summary Statistics for Inflammatory Bowel Disease

The IBD, CD, and UC summary statistics used in the present study are previously published and publicly accessible.[12] The complete methods have been published previously[12] but are presented here in brief. Following ethical approval by Cambridge MREC, 11 768 UK IBD cases determined by accepted endoscopic, histopathological, and radiological diagnostic criteria were consented and genotyped on the Human Core Exome v12.1 chip.[12] In total, 10 484 population control samples were obtained from the Understanding Society Project, and these samples were genotyped on the Human Core Exome v12.0 chip.[12] Genotypes were called using optiCall, and SNP-level quality control was performed by removing variants with the following criteria: not observed on both versions of the chips, missingness >5%, significant differences in call rates between cases and controls, deviated from Hardy-Weinberg equilibrium in controls, or affected by genotyping batch effect.[12] Sample-level quality control involved removing samples with missingness >1%, heterozygosity +/- 3 SD from the mean, mismatch between reported and genotyped sex, and non-European samples identified through principal component analysis (PCA) with HapMap3 populations.[12] After these quality control steps were completed, the final cohort included data for 4474 patients with CD, 4173 patients with UC, 592 patients with IBD-unclassified cases, and 9500 control samples directly genotyped at 296 203 variants.[12] The SNP imputation into these data was done from the HapMap3 reference panel, yielding approximately 1.1 million loci for association testing.[12] The association summary statistics used in the present study include all IBD cases for the IBD cohort (CD, UC, and IBD-unclassified cases), CD patients only for the CD cohort, and UC patients only for the UC cohort.

Behavioral, Diet-related, Psychiatric, and Clinical Trait Accession From United Kingdom Biobank Genome-Wide Studies

Behavioral, diet-related, psychiatric, and clinical traits used in the cross-trait LDSR analyses were obtained from the United Kingdom Biobank (UKBB). The UKBB is a large prospective study and health data repository with over 500 000 participants aged 40 to 69 years when recruited from 22 sites across the UK from 2006 to 2010.[20] The UKBB collected extensive phenotypic and genotypic detail about participants, including data from surveys, physical measurements, sample assays, accelerometry, multimodal imaging, genome-wide

genotyping, and longitudinal follow-up for a broad range of health-related outcomes.[20] At the time of recruitment, information on food consumption frequency was collected via a touchscreen questionnaire. Participants were asked to report their current (ie, at time of recruitment) intake frequency of a range of common food and drink items, ranging from vegetable intake to meat and dairy intake. The UKBB is open-access and available freely to researchers who seek to use it to conduct health-related research for public benefit.[20] Each of the over 500 000 participants in the UKBB has been genotyped, 90% of which used a custom Affymetrix UKBB Axiom array.[18,20] The array assayed approximately 850 000 SNPs across the genome, which were then used to impute 9.1 million SNPs through a collaboration with the Wellcome Trust Center for Human Genetics.[18,20] The GWAS was conducted using the imputed data, and summary level statistics are publicly available (http://nealelab.github.io/UKBB_ldsc/downloads.html#reference_files). The data used in this study were obtained from the second batch of UKBB GWAS results published online (updated August 2018).

Harmonization and Quality Control With SNP Filtering

Using an approach that we have previously published,[18] we harmonized the publicly available GWAS summary statistics. The final data set included summary statistics from the UKBB for behavioral, diet-related, psychiatric, and clinical traits that met specific confidence criteria, as outlined by the Neale lab. First, confidence in the h_g^2 (genetic heritability) is evaluated by determining the minimum effective sample size (N_{eff}) with a useful level of power to detect genetic heritability. Expected standard errors (SEs) are modeled as a function of sample size, which is then used for power analysis. Traits with a favorable $N_{eff} > 40 000$ are considered "high" confidence. In addition to selecting traits with "high" confidence, a P value can also be determined by comparing the expected distribution of h_g^2 to the actual distribution of h_g^2. Only traits with a highly significant P value ($P < 3.17 \times 10^{-5}$) as previously described[19] were included in our analysis as a quality control step. Each selected UKBB summary statistic contained SNP-level effect sizes (beta) for each trait, along with Z scores calculated by dividing beta by the standard error. Finally, as an additional quality control measure, we filtered the imputed SNPs from the UKBB to include only those autosomal SNPs with a minor allele frequency (MAF) greater than 1% and an imputation

quality INFO score greater than 0.90.[18] Additionally, SNPs that were not in HapMap3 with a minor allele frequency less than 5% in European populations were removed, which is in line with previously published methods.[18,21] All summary statistics were harmonized to a final format understandable by LDSR, which includes the variant-specific RS number, the effect and noneffect alleles, the sample size, a P value, and a signed summary statistic. In the final analysis, 1 045 095 common SNPs were included for IBD and the UKBB traits. Similarly, there were 1 045 070 common SNPs for CD and the UKBB traits and 1 045 057 common SNPs for UC and the UKBB traits.

Linkage Disequilibrium Score Regression to Estimate Genetic Correlations and Heritability

Using the harmonized summary statistics, we estimated genome-wide cross-trait genetic correlations between the UKBB traits with IBD, CD, and UC using LDSR (Figure 1). Linkage disequilibrium score regression leverages patterns of linkage disequilibrium (LD) between SNPs across the genome to infer a shared genetic basis between traits.[19,21] For each SNP in the genome, an LD score is calculated that captures the extent of pairwise LD between a given SNP and other SNPs in the genome. The LD scores were derived from the HapMap3 reference panel of individuals with known genotype information. Linkage disequilibrium score regression was then used to calculate the genetic correlation between 2 traits by regressing the product of SNP z scores ($Z_{UKBB} \times Z_{IBD}$) against the SNP's calculated LD score.[21] The slope of the regression estimates the genetic covariance between the 2 traits of interest, and this genetic covariance is converted to a genetic correlation value (r_g) by normalizing genetic covariance by the calculated heritability of the 2 compared traits.[19] The intercept term accounts for genomic inflation, serving to decrease biases that may arise from population stratification and cryptic relatedness.[21,22] In the present study, we utilized a cross-trait LDSR model that included an intercept to account for hidden biases that may arise due to the instability of LD scores in European subpopulations, as previously described.[18] Traits achieving at least nominal significance ($P < .05$) were analyzed in the present analysis, and Bonferroni correction for multiple comparisons (denoted as Bonferroni P value and calculated by multiplying the P value by the number of associations) is included in Supplementary Tables 1-3.
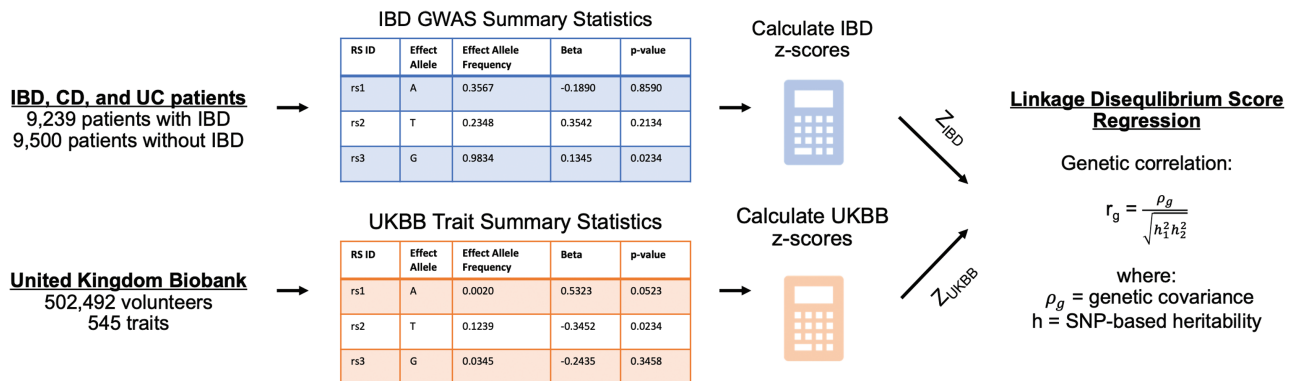
Figure 1. Graphical representation of analytical workflow to obtain genetic correlations (r_g) via LDSR. Note: Example data were used in the tables to illustrate the data structure. Figure adapted from Pettit et al. (2021).[18]

Results

Heritability of IBD, CD, and UC and Genetic Correlation Between CD and UC

Using LDSR to calculate SNP-based heritability (Figure 1), we found the heritability of IBD to be 29.6%, with a standard error of 2.6%. The estimated heritability of CD and UC were 41.8% ± 4.4% and 24.5% ± 2.3%, respectively (Figure 2). The estimated heritability of the 545 UKBB traits modeled using LDSR appear in Supplementary Tables 1 to 3. The pairwise genetic correlation between CD and UC is 0.62 (±0.03; $P = 1.51 \times 10^{-79}$).

Genetic Correlations Between IBD and Behavioral Traits

The cross-trait genetic correlation (r_g) was determined between IBD, CD, and UC and various traits related to smoking, alcohol use, lifestyle, and educational attainment (Figure 3). A P value (p_b) calculated by multiplying the original P value by the number of analyzed traits ($n = 545$) for traits reaching Bonferroni-corrected significance is included; traits without a P_b value did not reach Bonferroni-corrected significance. Current tobacco smoking and current smoking status were positively correlated with CD ($r_g = 0.12$, $P = 4.2 \times 10^{-4}$; $r_g = 0.11$, $P = 1.4 \times 10^{-3}$, respectively), while the smoking status "never" was negatively correlated with CD ($r_g = -0.09$, $P = 2.5 \times 10^{-3}$). The presence of smoke or smokers in the household and ever smoking were negatively correlated with UC ($r_g = -0.22$, $P = 4.7 \times 10^{-3}$; $r_g = -0.07$, $P = .042$, respectively). Current alcohol drinking was negatively correlated with IBD and CD ($r_g = -0.13$, $P = 2.4 \times 10^{-3}$; $r_g = -0.18$, $P = 5.0 \times 10^{-5}$, $P_b = 0.027$, respectively), while previous alcohol drinking was positively correlated with IBD and CD ($r_g = 0.15$, $P = 5.4 \times 10^{-3}$; $r_g = 0.18$, $P = 9.2 \times 10^{-4}$, respectively).

Next, we investigated lifestyle and educational attainment traits. Traits such as time spent watching television (TV) and time spent outdoors in summer were positively correlated with CD ($r_g = 0.09$, $P = 3.0 \times 10^{-3}$; $r_g = 0.09$, $P = 8.1 \times 10^{-3}$, respectively). The self-reported educational attainment traits come from the UKBB, meaning educational attainment metrics follow the United Kingdom's schooling schema. Professional qualifications (eg, nursing, teaching) were negatively correlated with IBD and CD ($r_g = -0.14$, $P = 1.3 \times 10^{-4}$; $r_g = -0.16$, $P = 3.1 \times 10^{-6}$, $P_b = 1.7 \times 10^{-3}$, respectively). Additionally, obtaining a college/university degree was negatively correlated with IBD and CD ($r_g = -0.11$, $P = 2.3 \times 10^{-4}$; $r_g = -0.15$, $P = 2.4 \times 10^{-8}$, $P_b = 1.3 \times 10^{-5}$, respectively). Fluid intelligence score was negatively correlated with IBD, CD, and UC ($r_g = -0.11$, $P = 5.3 \times 10^{-4}$; $r_g = -0.12$, $P = 3.6 \times 10^{-4}$; $r_g = -0.9$, $P = .011$, respectively). All traits with at least 1 nominally significant ($P < .05$) genetic correlation are shown in Figure 3 and the absolute value of differences in the genetic correlation between CD and UC are shown in Supplementary Figure 1.
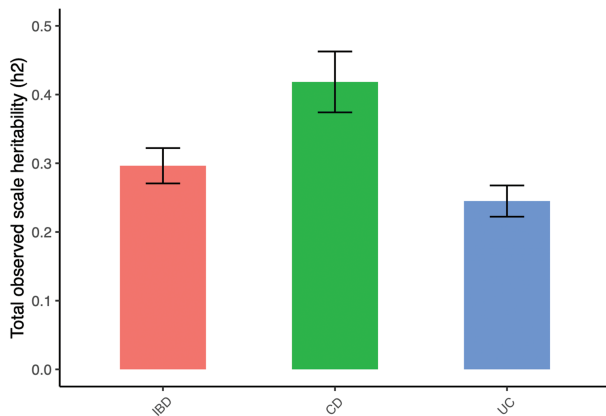

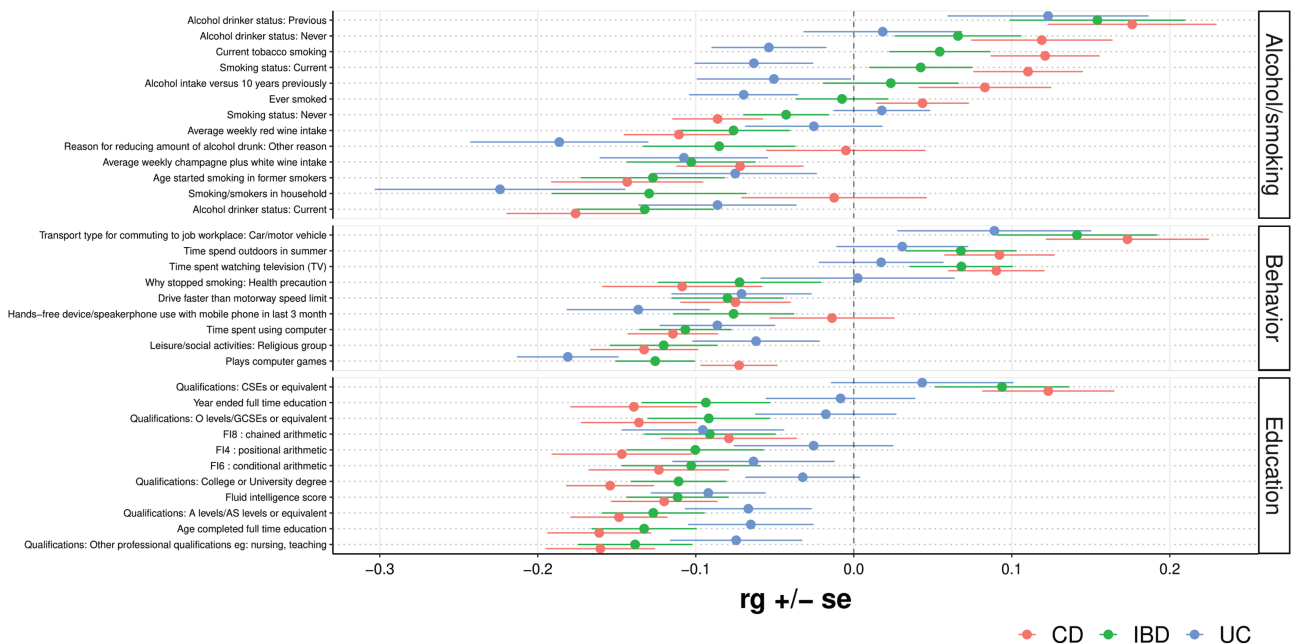**Figure 2.** Total observed scale heritability (h2) for IBD, CD, and UC.

Figure 3. Forest plot showing genetic correlations (r_g) for alcohol/smoking, behavior, and education related traits. Note: Only traits with at least 1 nominally significant ($P < .05$) genetic correlation are shown.

Genetic Correlations Between IBD and Diet-related Traits

Next, the cross-trait r_g was determined between IBD, CD, and UC and several diet-related traits (Figure 4). Major dietary changes in the last 5 years because of illness were positively correlated with IBD, CD, and UC ($r_g = 0.19$, $P = 1.1 \times 10^{-6}$, $P_b = 6.1 \times 10^{-4}$; $r_g = 0.19$, $P = 4.2 \times 10^{-7}$, $P_b = 2.3 \times 10^{-4}$; $r_g = 0.15$, $P = 1.6 \times 10^{-3}$, respectively). Bread intake was positively correlated with UC ($r_g = 0.09$, $P = 8.3 \times 10^{-3}$). Cereal type of "other (e.g., Cornflakes, Frosties)," white bread type, salt added to food, and soft tub margarine nonbutter spread type were all positively correlated with CD, while muesli cereal type, whole meal or wholegrain bread type, ground coffee type, fresh fruit intake, fish oil use, salad and raw vegetable intake, and coffee intake were all negatively correlated with CD. Dried fruit intake and fresh fruit intake were negatively correlated with IBD ($r_g = -0.11$, $P = 3.5 \times 10^{-5}$, $P_b = .019$; $r_g = -0.06$, $P = .044$, respectively). All traits with at least 1 nominally significant ($P < .05$) genetic correlation are shown in Figure 4.

Genetic Correlations Between IBD and Psychiatric Traits

We next analyzed the cross-trait r_g between IBD, CD, and UC and several psychiatric-related traits (Figure 5). Globally, a strong qualitative genetic correlation was observed across IBD, CD, and UC for most of the analyzed traits, many of which focus on traits related to anxiety and depression. A strong positive genetic correlation was observed for a trait describing recent easy annoyance or irritability for IBD, CD, and UC ($r_g = 0.21$, $P = 2.0 \times 10^{-4}$; $r_g = 0.20$, $P = 1.6 \times 10^{-4}$; $r_g = 0.18$, $P = 9.3 \times 10^{-3}$, respectively). A strong positive genetic correlation was also observed for a trait describing being tense or "highly strung" with IBD, CD, and UC ($r_g = 0.15$, $P = 1.2 \times 10^{-6}$, $P_b = 6.8 \times 10^{-4}$; $r_g = 0.09$,

$P = 2.8 \times 10^{-3}$; $r_g = 0.18$, $P = 2.2 \times 10^{-6}$, $P_b = 1.2 \times 10^{-3}$, respectively). Furthermore, a positive genetic correlation was observed between the trait "seen a psychiatrist for nerves, anxiety, tension or depression" and IBD, CD, and UC ($r_g = 0.13$, $P = 1.6 \times 10^{-4}$; $r_g = 0.14$, $P = 1.6 \times 10^{-5}$, $P_b = 8.7 \times 10^{-3}$; $r_g = 0.10$, $P = .016$, respectively). Additional traits with a positive genetic correlation with IBD, CD, and UC include worrying more than others in similar situations, seeing a general practitioner for nerves, anxiety, tension, or depression, prolonged feelings of sadness or depression, mood swings, and unenthusiasm/disinterest in the last 2 weeks, among several other traits. All traits with at least 1 nominally significant ($P < .05$) genetic correlation are shown in Figure 5.

Genetic Correlations Between IBD and General Wellness Traits

Next, we analyzed the cross-trait r_g between IBD, CD, and UC and various traits related to general wellness, which broadly includes traits that did not fit into the other categories (Figure 6). A positive correlation was observed between IBD, CD, and UC and bowel cancer screening ($r_g = 0.28$, $P = 2.0 \times 10^{-8}$, $P_b = 1.1 \times 10^{-5}$; $r_g = 0.21$, $P = 2.8 \times 10^{-5}$, $P_b = .015$; $r_g = 0.30$, $P = 5.7 \times 10^{-7}$, $P_b = 3.1 \times 10^{-4}$, respectively). Significant positive correlations were also observed for IBD, CD, and UC with traits related to work status, such as "attendance/disability/mobility allowance: disability living allowance" and "current employment status: unable to work because of sickness of disability." A negative correlation was observed for IBD and CD with the trait "types of physical activity in last 4 weeks: strenuous sports" ($r_g = -0.14$, $P = 6.1 \times 10^{-3}$; $r_g = -0.15$, $P = 9.2 \times 10^{-4}$, respectively). A significant negative correlation was also observed for usual walking pace and the trait "types of physical activity last 4 weeks: other exercises (eg: swimming, cycling, keep fit, bowling)" for IBD, CD, and
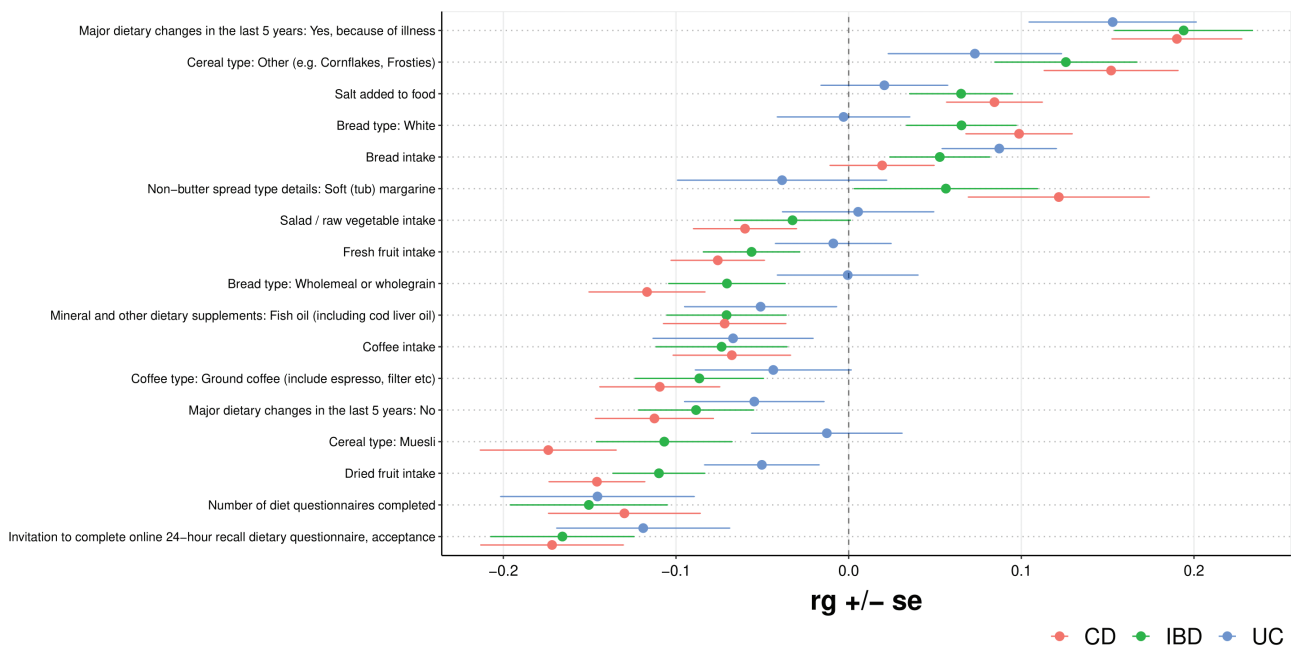
Figure 4. Forest plot showing genetic correlations (r_g) for diet-related traits. Note: Only traits with at least 1 nominally significant ($P < .05$) genetic correlation are shown.
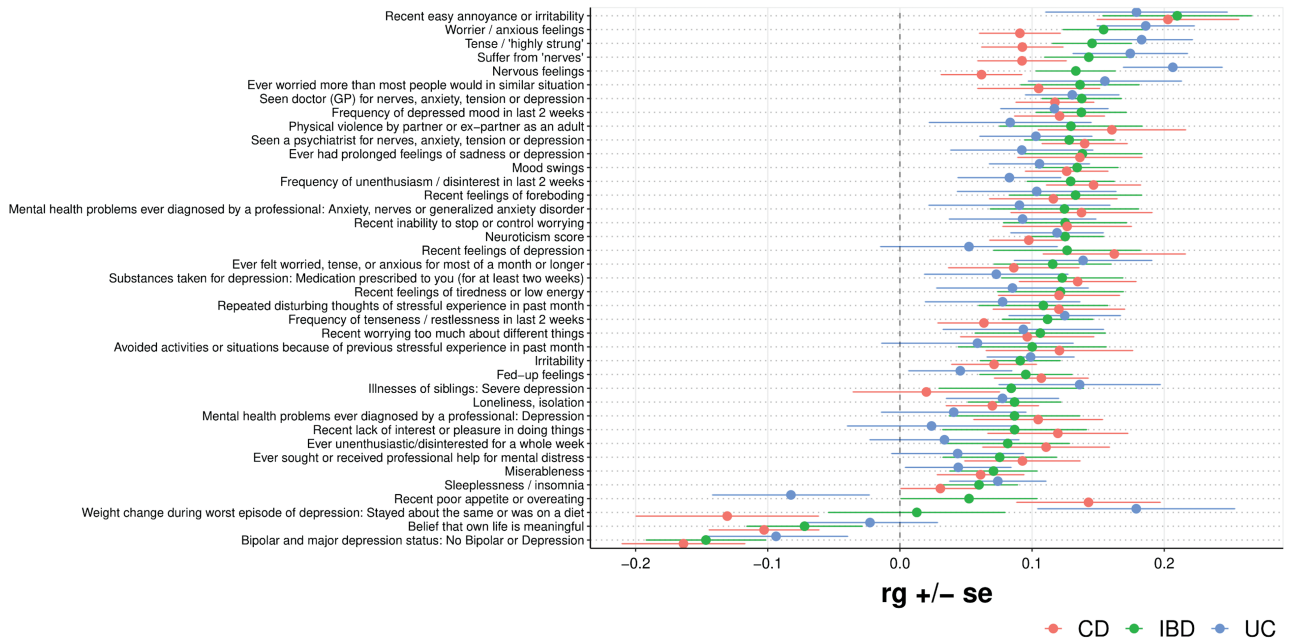
Figure 5. Forest plot showing genetic correlations (r_g) for psychiatric traits. Note: Only traits with at least 1 nominally significant ($P < .05$) genetic correlation are shown.

UC (except latter trait in UC). Interestingly, a positive genetic correlation was observed between BMI at time of UKBB assessment and CD ($r_g = 0.08$, $P = 3.6 \times 10^{-4}$), while a negative genetic correlation was observed between BMI at time of UKBB assessment and UC ($r_g = -0.07$, $P = .017$). All traits with at least 1 nominally significant ($P < .05$) genetic correlation are shown in Figure 6.

## Genetic Correlations Between IBD and Other Diseases and Physiological Traits

Finally, we analyzed the cross-trait r_g between IBD, CD, and UC and physician-diagnosed conditions or conditions and laboratory values present in patient electronic health records (EHRs; Figure 7). A strong positive correlation was observed for the trait "diseases of the digestive system" for IBD, CD, and UC ($r_g = 0.23$, $P = 7.8 \times 10^{-8}$, $P_b = 4.3 \times 10^{-5}$; $r_g = 0.23$, $P = 1.6 \times 10^{-8}$, $P_b = 6.3 \times 10^{-6}$; $r_g = 0.20$, $P = 4.2 \times 10^{-4}$ respectively). A positive correlation was also observed for unspecified hematuria and IBD and CD ($r_g = 0.21$, $P = 4.4 \times 10^{-3}$; $r_g = 0.22$, $P = 3.2 \times 10^{-3}$ respectively). A strong positive correlation was observed for pain in the throat and chest and IBD, CD, and UC ($r_g = 0.19$, $P = 5.9 \times 10^{-6}$, $P_b = 3.2 \times 10^{-3}$; $r_g = 0.20$, $P = 8.9 \times 10^{-7}$, $P_b = 4.9 \times 10^{-4}$; $r_g = 0.12$, $P = .019$, respectively). A positive correlation was observed between diagnosed hay fever or allergic rhinitis and CD ($r_g = 0.14$, $P = 2.8 \times 10^{-3}$). In addition to medical conditions diagnosed by physicians or extracted from patient EHR data, various laboratory values were analyzed. A positive genetic correlation was observed between CD and neutrophil percentage and neutrophil count ($r_g = 0.10$, $P = 9.2 \times 10^{-3}$; $r_g = 0.11$, $P = .020$, respectively), while a negative genetic correlation was observed between IBD and CD with basophil percentage ($r_g = -0.08$, $P = .010$; $r_g = -0.09$, $P = 9.2 \times 10^{-3}$, respectively). A negative genetic correlation was also observed between IBD and CD with lymphocyte percentage ($r_g = -0.09$, $P = .026$; $r_g = -0.10$, $P = 6.8 \times 10^{-3}$). All traits with at least 1

nominally significant ($P < .05$) genetic correlation are shown in Figure 7.

Discussion

In the present study, we provide an atlas of the shared genetic architecture of IBD, CD, and UC with various behavioral, diet-related, psychiatric, general wellness, and clinical traits. Linkage disequilibrium score regression has been previously used to determine the genetic correlation and shared heritability between traits (phenotypes) and diseases of interest both in other conditions (eg, lung cancer[18]) and between IBD and traits such as multiple sclerosis.[23] In the present study, we perform wide-scale cross-trait analysis between IBD, CD, and UC and hundreds of traits from the UKBB, finding many significant associations that may provide insight into IBD development, progression, and comorbidities.

Our study provides additional evidence that IBD and its main 2 subtypes, CD and UC, are highly heritable diseases. Up to 12% of IBD patients have a family history of IBD, and family and twin studies have demonstrated that host genetics play a larger role in CD development compared with UC.[24] Additionally, first-degree relatives of patients with CD and UC have an incidence rate ratio of 7.77 and 4.08, respectively.[9,24] The results from our study provide additional evidence that both major IBD subtypes are highly heritable, as we demonstrate an overall IBD heritability of 29.6% ± 2.6%, a CD heritability of 41.8% ± 4.4%, and a UC heritability of 24.5% ± 2.3%. These results support existing evidence that host genetics play a greater role in CD development, with our study demonstrating a 1.7-fold higher heritability in CD compared with UC. Notably, however, the SNP-based heritability described in our study differs from heritability established from family and twin studies for a few reasons, such as the assumption that SNP-based heritability assumes the effect of each SNP is independent and
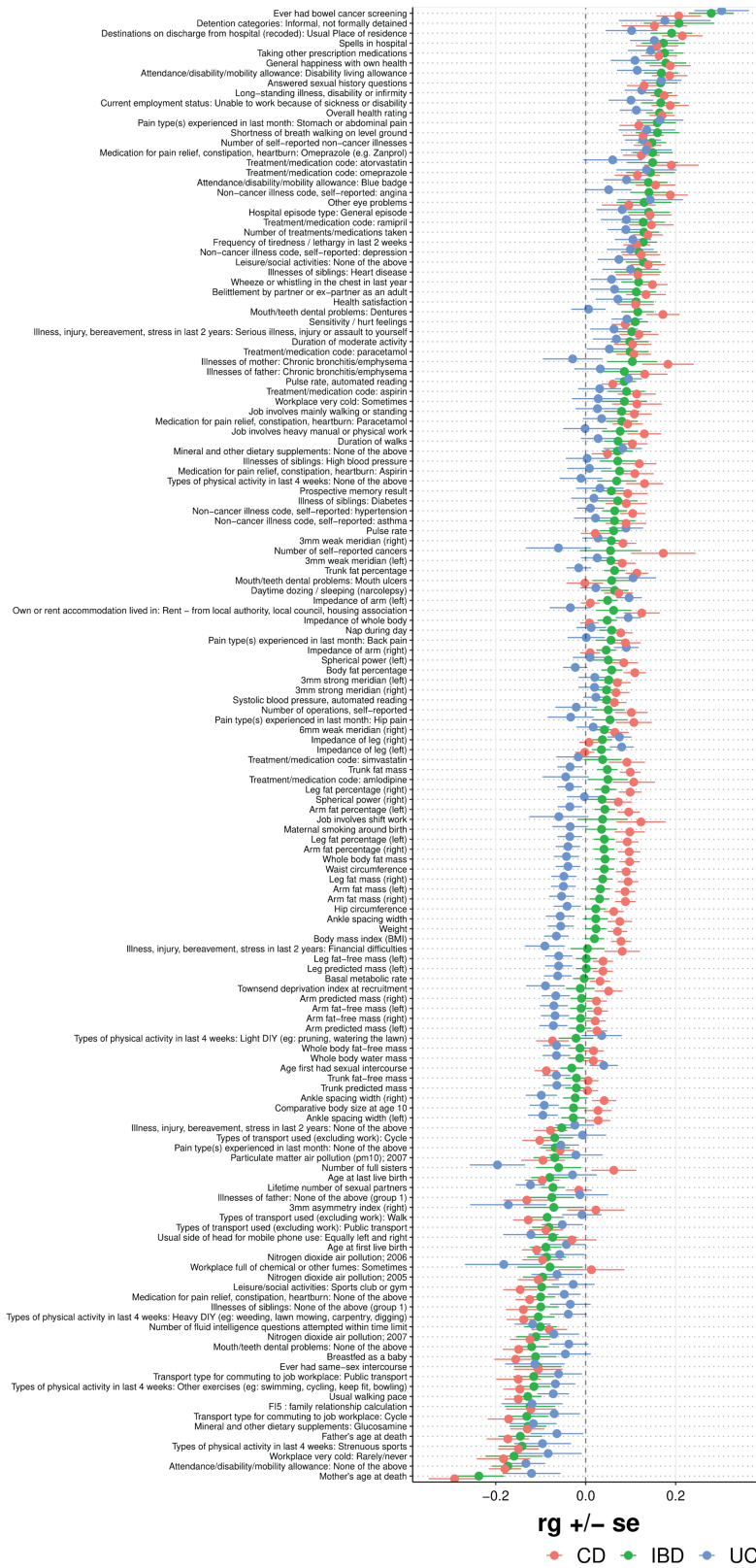
Figure 6. Forest plot showing genetic correlations (r_g) for general wellness traits. Note: Only traits with at least 1 nominally significant ($P < .05$) genetic correlation are shown.

additive. This assumption may not hold true for all genetic variants. Additionally, SNP-based heritability focuses on additive effects of individual SNPs across the genome, while heritability from family and twin studies includes both additive

and nonadditive genetic effects, rare genetic variants, and gene-environment interactions.

One particular area of inquiry in IBD has been the differential impact of smoking behavior on CD vs UC. Early studies
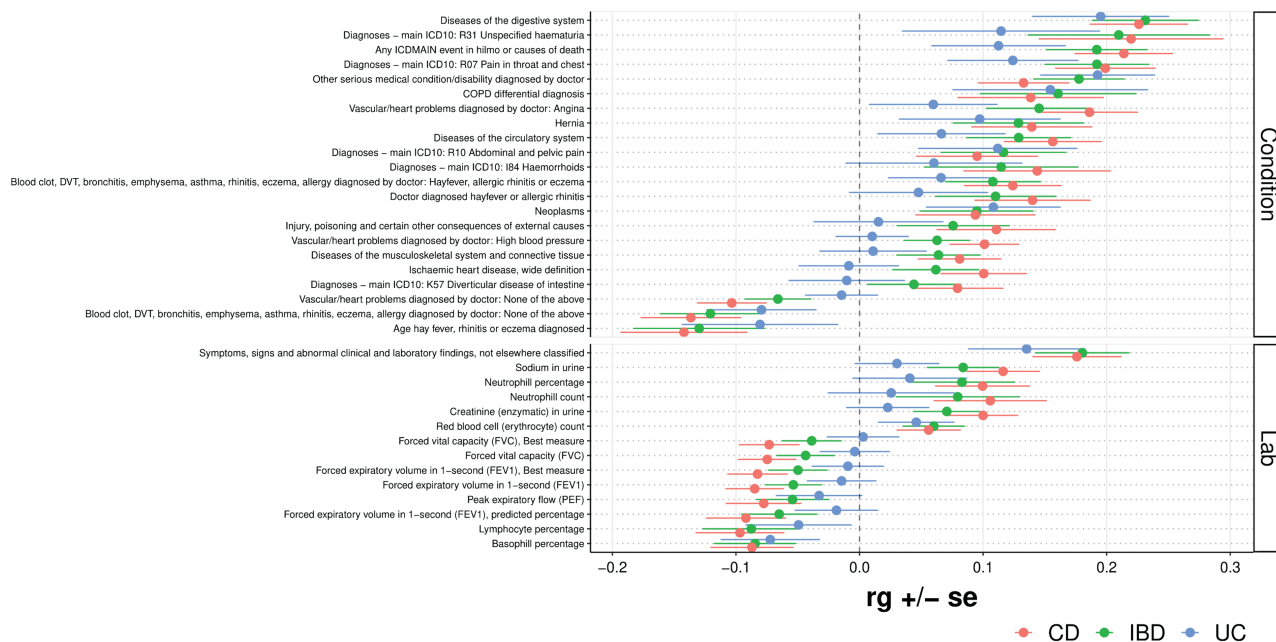
Figure 7. Forest plot showing genetic correlations (r_g) for other clinical conditions or laboratory traits. Note: Only traits with at least 1 nominally significant ($P < .05$) genetic correlation are shown.

showed a reduced frequency of smokers among patients with UC compared with healthy controls, while a meta-analysis in CD demonstrated an increased risk associated with smoking (odds ratio [OR], 1.76, 95% confidence interval [CI], 1.40-2.22).[15,25,26] A similar association was seen between *former* smoking at the time of UKBB assessment and UC—but not *current* smoking at the time of UKBB assessment and UC, as the latter demonstrates a strong inverse association (OR, 0.58, 95% CI, 0.45-0.75).[25] Our study demonstrates a similar pattern between the cross-trait genetic correlations of smoking traits with CD and UC. Specifically, current tobacco smoking and current smoking status traits were positively correlated with CD, while never smoking was negatively genetically correlated with CD. On the other hand, a negative (ie, protective) genetic correlation was seen between the presence of smoke or smokers in the household and ever smoking with UC. The genetic correlations reached nominal significance, but they did not reach Bonferroni-corrected significance. Various hypotheses have been set forth to explain the effect of smoking on IBD, such as the effect of smoking on smooth muscle tone and endothelial function through nitric oxide production, an effect on the integrity of the gut mucosal barrier, an increase in oxidative effects, an influence on the gut microbiota, or a differential production of mononuclear cells in CD but not UC in response to smoke.[15]

Education-related traits demonstrated interesting and consistent negative genetic correlations with IBD, CD, and UC in the LDSR analyses. Professional qualifications (eg, nursing, teaching) were negatively correlated with IBD and CD, and obtaining a college/university degree was negatively correlated with IBD and CD. The aforementioned correlations involving CD achieved Bonferroni-corrected significance. Additionally, a nominally significant negative genetic correlation was observed between IBD, CD, and UC and fluid intelligence score, described in the UKBB as the capacity to solve problems that require logic and reasoning ability, independent of acquired knowledge. The genetic relationship between educational attainment

and IBD, CD, and UC is poorly characterized. Studies have, however, demonstrated poorer outcomes in IBD patients with lower socioeconomic status (SES), and educational attainment may be a proxy variable for SES.[27] For example, one study demonstrated a 29% increase of in-hospital mortality for Crohn's disease patients with a median income below the average.[28] Another study found a 3.3-fold increased odds of in-hospital mortality for UC patients with Medicaid coverage, a form of insurance provided for low-income patients in the United States.[29] Taken together, these results may suggest a compounding effect between educational attainment, income, and IBD, and future studies should seek to disentangle the temporal relationship between these variables.

The role of diet in IBD development and management is another well-studied topic, and nutritional research groups have developed guideline diets for patients with IBD, such as the Crohn's Disease Exclusion Diet[30] or the Mediterranean diet. In the present study, we demonstrated an expected strong positive genetic correlation between major dietary changes in the last 5 years due to illness with IBD, CD, and UC, with the IBD and CD correlations achieving Bonferroni-corrected significance. While fewer significant genetic correlations were observed for UC, which has a lower genetic heritability than CD, several significant genetic correlations were observed for CD. For example, white bread type, salt added to food, and soft tub margarine nonbutter spread type were all positively correlated with CD while muesli cereal type, whole meal or wholegrain bread type, ground coffee type, fresh fruit intake, fish oil use, salad and raw vegetable intake, and coffee intake were all negatively correlated with CD. Many raw vegetables are high in vitamin D and/or fiber, and epidemiologic studies suggest that vitamin D and fiber intake may be protective in CD.[15] Data from the Nurses' Health Study, a longitudinal study following 170 776 women for 26 years, suggest that long-term intake of dietary fiber, especially from fruit, is associated with a lower risk of CD but not UC, which supports the findings found in the present study.[31]

The psychiatric burden associated with IBD has been well-documented, and a meta-analysis robustly demonstrated a lower quality of life (QoL) for individuals with IBD compared with healthy individuals, for both adults and children.[32] In the present study, we observed both nominal and Bonferroni-corrected significant genetic correlations between IBD, CD, and UC with most of the analyzed traits, many of which are survey-based questions related to anxiety and depression. Strongly positively correlated traits included seeing a psychiatrist for a mental health condition and prolonged feelings of sadness or depression. An interesting negative genetic correlation was observed between IBD and CD and not having bipolar disorder, suggesting that there may exist a positive genetic correlation between the diseases. Studies suggest that the relationship between IBD and anxiety and depression is bidirectional, and the mechanisms underlying this relationship may include increased pro-inflammatory cytokines, vagal nerve signaling, gut dysbiosis, and changes in brain signaling and morphology.[2] Additionally, one study found that symptoms of anxiety and depression were associated with more aggressive IBD.[33] Taken together with the results from the present study, it is clear that IBD and psychiatric conditions, such as anxiety and depression, are highly comorbid. The approach used in this study compares genetic factors influencing IBD to those influencing psychiatric and behavioral factors, hence these correlations exist prior to IBD development. However, the extent to which genetic determinants of subclinical or undiagnosed IBD may also contribute to the observed polygenicity of neuropsychiatric conditions remains unknown and merits future research.

Finally, the cross-trait genetic correlation was analyzed between IBD, CD, and UC and an array of physician-diagnosed and laboratory values. In addition to a Bonferroni-significant positive genetic correlation between traits associated with IBD, such as diseases of the digestive system or pain in the throat and chest, a nominal positive genetic correlation was observed between IBD and CD with unspecified hematuria. A retrospective study found that children with IBD are more likely to show kidney-related symptoms, such as isolated hematuria and/or proteinuria than healthy children and adolescents.[34] Renal and urinary involvement is estimated to occur in 4% to 23% of patients with IBD, with a significant effect on overall patient morbidity.[35] Optimal screening guidelines for renal function monitoring in patients with IBD have not yet been established, but taken together, these findings suggest that renal function monitoring may be an important aspect of IBD management. Furthermore, in the present study, a positive genetic correlation was observed between diagnosed hayfever or allergic rhinitis, neutrophil percentage, and neutrophil count with CD. Studies have previously demonstrated an epidemiologic association between hay fever and IBD.[36] Evidence also suggests that neutrophils play a role in IBD development and/or management by forming neutrophil extracellular trap (NET) that can potentiate gut inflammation and mediate IBD-related complications.[37] Taken together, the genetic correlations validate known associations and provide additional evidence that advances our understanding of the etiology and comorbidities associated with IBD.

The present study has several limitations. First, we do not establish the causal directionality between IBD and the UKBB traits; the results are genetic associations, not causal effects. The potential for confounding or mediating traits that contribute to the observed effects is possible for the observed genetic correlations. Additionally, many UKBB traits are presented as binary traits, while the participants may have responded to ordinal surveys potentially removing nuance from participant responses. Another data limitation is that terms from patient electronic health records, such as "throat/chest pain," are nonspecific. Additionally, diet-related questions may fail to capture temporal changes in participant diets. Finally, the present analysis included only European patients due to data availability, limiting extension of the current results to diverse populations. For traits with interesting associations, performing Mendelian randomization (MR) analysis will be an important next step. Notably, however, LDSR uses genome-wide data, while MR uses selected causal markers, making the latter less powerful for discovery.

Our current study presents an atlas of cross-trait genetic correlations between IBD, CD, and UC with hundreds of important behavioral, psychiatric, diet-related, general wellness, and clinical traits from the UKBB. Our work provides genetic confirmation of many epidemiologically established relationships, serving as a step towards understanding the shared genetic architecture between IBD, CD, and UC with these traits. Future partitioned heritability analyses may be used to investigate enrichment of IBD heritability within specific immune cell subsets. Identifying causal vs outcome relationships may help to guide treatment decisions, update diet or behavioral recommendations, and generate more accurate predictive risk models. Furthermore, the significant genetic relationships established here may be used as the input for a loci-trait unsupervised clustering method for deconstructing the germline genetic heterogeneity of IBD that may underlie variable disease course and response to treatments.

Supplementary Data

Supplementary data is available at *Inflammatory Bowel Diseases* online.

Acknowledgments

Funding

Conflicts of Interest

The authors declare no potential conflicts of interest.

Data Availability

The IBD GWAS data are accessible via the European Bioinformatics Institute GWAS Catalog (https://www.ebi.ac.uk/gwas/publications/28067908). The UKBB GWAS data are accessible via the UKBB website (https://biobank.ndph.ox.ac.uk/ukb/exinfo.cgi?src=AccessingData).

References

1. 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. doi:10.1016/S0140-6736(17)32448-0

2. Bisgaard TH, Allin KH, Keefer L, Ananthakrishnan AN, Jess T. Depression and anxiety in inflammatory bowel disease: epidemiology, mechanisms and treatment. *Nat Rev Gastroenterol Hepatol.* 2022;19(11):717-726. doi:10.1038/s41575-022-00634-6

3. Kuenzig ME, Manuel DG, Donelle J, Benchimol EI. Life expectancy and health-adjusted life expectancy in people with inflammatory bowel disease. *CMAJ.* 2020;192(45):E1394-E1402. doi:10.1503/cmaj.190976

4. Peyrin-Biroulet L, Panés J, Sandborn WJ, et al. Defining disease severity in inflammatory bowel diseases: current and future directions. *Clin Gastroenterol Hepatol.* 2016;14(3):348-354.e17. doi:10.1016/j.cgh.2015.06.001

5. Annese V. Genetics and epigenetics of IBD. *Pharmacol Res.* 2020;159:104892. doi:10.1016/j.phrs.2020.104892

6. Sudhakar P, Verstockt B, Cremer J, et al. Understanding the molecular drivers of disease heterogeneity in Crohn's disease using multi-omic data integration and network analysis. *Inflamm Bowel Dis.* 2021;27(6):870-886. doi:10.1093/ibd/izaa281

7. McGovern DPB, Kugathasan S, Cho JH. Genetics of inflammatory bowel diseases. *Gastroenterology.* 2015;149(5):1163-1176.e2. doi:10.1053/j.gastro.2015.08.001

8. 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. doi:10.1038/nrgastro.2017.136

9. Moller FT, Andersen V, Wohlfahrt J, Jess T. Familial risk of inflammatory bowel disease: a population-based cohort study 1977-2011. *Am J Gastroenterol.* 2015;110(4):564-571. doi:10.1038/ajg.2015.50

10. Yamazaki K, McGovern D, Ragoussis J, et al. Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum Mol Genet.* 2005;14(22):3499-3506. doi:10.1093/hmg/ddi379

11. Luo Y, de Lange KM, Jostins L, et al. Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. *Nat Genet.* 2017;49(2):186-192. doi:10.1038/ng.3761

12. 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. doi:10.1038/ng.3760

13. Liu JZ, van Sommeren S, Huang H, et al.; International Multiple Sclerosis Genetics Consortium. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet.* 2015;47(9):979-986. doi:10.1038/ng.3359

14. Geremia A, Biancheri P, Allan P, Corazza GR, Di Sabatino A. Innate and adaptive immunity in inflammatory bowel disease. *Autoimmun Rev.* 2014;13(1):3-10. doi:10.1016/j.autrev.2013.06.004

15. Ananthakrishnan AN. Epidemiology and risk factors for IBD. *Nat Rev Gastroenterol Hepatol.* 2015;12(4):205-217. doi:10.1038/nrgastro.2015.34

16. Raman M, Ghosh S. Diet and nutrition in IBD-progress and gaps. *Nutrients.* 2019;11(8):1740. doi:10.3390/nu11081740

17. Spencer EA, Agrawal M, Jess T. Prognostication in inflammatory bowel disease. *Front Med.* 2022;9:1025375. doi:10.3389/fmed.2022.1025375

18. Pettit RW, Byun J, Han Y, et al. The shared genetic architecture between epidemiological and behavioral traits with lung cancer. *Sci Rep.* 2021;11(1):17559. doi:10.1038/s41598-021-96685-x

19. Bulik-Sullivan B, Finucane HK, Anttila V, et al.; ReproGen Consortium. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015;47(11):1236-1241. doi:10.1038/ng.3406

20. 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. doi:10.1371/journal.pmed.1001779

21. Bulik-Sullivan BK, Loh PR, Finucane HK, et al.; Schizophrenia Working Group of the Psychiatric Genomics Consortium. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291-295. doi:10.1038/ng.3211

22. Yang J, Weedon MN, Purcell S, et al.; GIANT Consortium. Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet.* 2011;19(7):807-812. doi:10.1038/ejhg.2011.39

23. Yang Y, Musco H, Simpson-Yap S, et al. Investigating the shared genetic architecture between multiple sclerosis and inflammatory bowel diseases. *Nat Commun.* 2021;12(1):5641. doi:10.1038/s41467-021-25768-0

24. Turpin W, Goethel A, Bedrani L, Croitoru MK. Determinants of IBD heritability: genes, bugs, and more. *Inflamm Bowel Dis.* 2018;24(6):1133-1148. doi:10.1093/ibd/izy085

25. Mahid SS, Minor KS, Soto RE, Hornung CA, Galandiuk S. Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clin Proc.* 2006;81(11):1462-1471. doi:10.4065/81.11.1462

26. Harries AD, Baird A, Rhodes J. Non-smoking: a feature of ulcerative colitis. *Br Med J (Clin Res Ed).* 1982;284(6317):706. doi:10.1136/bmj.284.6317.706

27. Sewell JL, Velayos FS. Systematic review: the role of race and socioeconomic factors on IBD healthcare delivery and effectiveness. *Inflamm Bowel Dis.* 2013;19(3):627-643. doi:10.1002/ibd.22986

28. Nguyen GC, Bayless TM, Powe NR, Laveist TA, Brant SR. Race and health insurance are predictors of hospitalized Crohn's disease patients undergoing bowel resection. *Inflamm Bowel Dis.* 2007;13(11):1408-1416. doi:10.1002/ibd.20200

29. Nguyen GC, Laveist TA, Gearhart S, Bayless TM, Brant SR. Racial and geographic variations in colectomy rates among hospitalized ulcerative colitis patients. *Clin Gastroenterol Hepatol.* 2006;4(12):1507-1513. doi:10.1016/j.cgh.2006.09.026

30. Levine A, Wine E, Assa A, et al. Crohn's disease exclusion diet plus partial enteral nutrition induces sustained remission in a randomized controlled trial. *Gastroenterology.* 2019;157(2):440-450.e8. doi:10.1053/j.gastro.2019.04.021

31. Ananthakrishnan AN, Khalili H, Konijeti GG, et al. A prospective study of long-term intake of dietary fiber and risk of Crohn's disease and ulcerative colitis. *Gastroenterology.* 2013;145(5):970-977. doi:10.1053/j.gastro.2013.07.050

32. Knowles SR, Graff LA, Wilding H, et al. Quality of life in inflammatory bowel disease: a systematic review and meta-analyses-part I. *Inflamm Bowel Dis.* 2018;24(4):742-751. doi:10.1093/ibd/izx100

33. Gao X, Tang Y, Lei N, et al. Symptoms of anxiety/depression is associated with more aggressive inflammatory bowel disease. *Sci Rep.* 2021;11(1):1440. doi:10.1038/s41598-021-81213-8

34. Jang HM, Baek HS, Kim JE, et al. Renal involvement in children and adolescents with inflammatory bowel disease. *Korean J Pediatr.* 2018;61(10):327-331. doi:10.3345/kjp.2018.06485

35. Ambruzs JM, Larsen CP. Renal manifestations of inflammatory bowel disease. *Rheum Dis Clin North Am.* 2018;44(4):699-714. doi:10.1016/j.rdc.2018.06.007

36. Card TR, Langan SM, Chu TPC. Extra-gastrointestinal manifestations of inflammatory bowel disease may be less common than previously reported. *Dig Dis Sci.* 2016;61(9):2619-2626. doi:10.1007/s10620-016-4195-1

37. Drury B, Hardisty G, Gray RD, Ho GT. Neutrophil extracellular traps in inflammatory bowel disease: pathogenic mechanisms and clinical translation. *Cell Mol Gastroenterol Hepatol.* 2021;12(1):321-333. doi:10.1016/j.jcmgh.2021.03.002