**Multi-ancestry genome-wide and transcriptome-wide association analyses identified new risk loci and genes for inflammatory bowel disease**

Linshuoshuo Lyu[1†], Qing Li[1†], Chao Li[1], Ghadeer K Dawwas[1], You Chen[2], Ran Tao[3,4], Qi Liu[3], Wanqing Wen[1], Xiao-Ou Shu[1], May Washington[5], David A Schwartz[6], Wei Zheng[1], Zhijun Yin[2], Ken S Lau[7,8], Xingyi Guo[1,2*]

Affiliations:
[1] Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA.
[2] Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville TN 37203, USA.
[3] Department of Biostatistics, Vanderbilt University Medical Center, Nashville TN 37203, USA.
[4] Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37203, USA
[5] Department of Pathology, Vanderbilt University Medical Center, Vanderbilt University School of Medicine, Nashville TN 37203, USA.
[6] Division of Gastroenterology, Hepatology, and Nutrition, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
[7] Epithelial Biology Center and Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, TN, USA.
[8] Center for Computational Systems Biology, Vanderbilt University, Nashville, TN, USA.

[†] Shared co-first authorship
* Corresponding author contact information:
Dr. Xingyi Guo
Division of Epidemiology, Department of Medicine
Vanderbilt Epidemiology Center, and Vanderbilt-Ingram Cancer Center,
Vanderbilt University School of Medicine
2525 West End Ave. Suite 330, Nashville, TN 37203
Tel: +1 615-936-3471; Fax: +1 615-332-0502
Email: xingyi.guo@vumc.org

**ABSTRACT**
To advance genetic understanding of inflammatory bowel disease (IBD), we conducted genome-wide association meta-analyses of 63,415 IBD cases of European and East Asian descendants and identified 90 previously unknown risk loci. Integrating multi-ancestry transcriptome-wide association studies (TWAS), cell type–specific TWAS, alternative splicing (AS-WAS), and alternative polyadenylation (APA-WAS) analyses using RNA-seq data from normal colon tissues of 707 European and 364 East Asian individuals, we uncovered 506 high-confidence IBD risk genes, including 384 not previously reported. These genes converge on immune regulation, microbial interaction, and other pathways central to IBD pathogenesis, with over half showing transcriptional dysregulation supported by single-cell and spatial omics analyses. Notably, 46 risk genes are targeted by 225 drugs that have been approved or in Phase II/III trials, including sulfasalazine already used in IBD therapy. Our study findings deepen the understanding of IBD genetics and support the development of precision medicine for its prevention and treatment.

**KEYWORDS**
Inflammatory Bowel Disease, GWAS, TWAS, single-cell RNA sequencing, risk loci, susceptibility genes, druggable targets

## INTRODUCTION

Inflammatory Bowel Disease (IBD) comprises a group of chronic, relapsing-remitting inflammatory conditions of the gastrointestinal tract, primarily Crohn's disease (CD) and ulcerative colitis (UC). Since 1900, the incidence of IBD has been rising strikingly in newly industrialized countries[1].The age- and sex-adjusted incidence of IBD was estimated at 10.9 per 100,000 person-years (95% confidence interval (CI): 10.6–11.2)[2]. Genome-wide association studies (GWAS) have identified more than 300 risk loci associated with IBD, primarily through large-scale studies in individuals of European and East Asian ancestry[3-10]. Recent efforts, including the largest East Asian IBD GWAS (14,393 cases and 15,456 controls)[4] and data from the FinnGen consortium (10,960 cases and 489,388 controls)[11], further expanded the catalog of IBD risk loci. However, the causal genes and biological mechanisms underlying these associations remain largely unresolved. Moreover, previous GWAS were conducted in single ancestry groups, which limits cross-population generalizability and hinders a comprehensive understanding of disease biology. Multi-ancestry GWAS meta-analysis can enhance risk locus discovery and refine association signals by leveraging shared genetic architectures.

Transcriptome-wide association studies (TWAS) offer a powerful framework to bridge GWAS signals with gene expression, by using predictive models trained on expression quantitative trait loci (eQTLs) to identify genes whose genetically predicted expression is associated with disease risk[12,13]. However, conventional TWAS approaches are often limited by the inclusion of non-functional variants in prediction models, reducing power and increasing false discovery rates[14-16]. To overcome these limitations, we developed sTF-TWAS, a framework that prioritizes regulatory variants within transcription factor-occupied cis-regulatory elements (STFCREs), significantly improving the accuracy and sensitivity of TWAS gene discovery[17,18]. TWAS has also been extended to investigate genetically regulated post-transcriptional features such as alternative splicing (AS) and alternative polyadenylation (APA), offering additional insights into gene regulation and disease mechanisms[19]. Furthermore, single-cell RNA sequencing (scRNA-seq) enables high-resolution characterization of tissue composition and facilitates the construction of cell type-specific transcriptomic profiles[20], thereby improving the detection of cell type-specific genetic effects[21-26]. Integration of scRNA-seq data with computational deconvolution tools, such as CIBERSORTx[27], has demonstrated utility in uncovering cell type-specific genetic regulation in several complex diseases[21-23,28-30]. However, such approaches remain underutilized in IBD, and to date, only limited TWAS of IBD using bulk normal colon tissue has been reported[31,32] underscoring the need for high-resolution, multi-omics integrative analyses in this field.

To address these critical gaps, we conducted a multi-ancestry meta-analysis of GWAS summary statistics from 63,415 IBD cases of European and East Asian ancestry to advance the discovery of novel risk loci. We next integrated the results from the multi-ancestry TWAS in conducting cell type-specific TWAS, alternative splicing-wide association study (AS-WAS), and alternative polyadenylation-wide association study (APA-WAS). The cell type-specific TWAS was performed by leveraging single-cell RNA-seq (scRNA-seq) reference data from Colon Molecular Atlas Project (COLON MAP)[33], and bulk RNA-seq data from 707 European individuals[34] and 364 East Asian[34]. We aimed to combine genetically predicted gene expression, AS, and APA with fine-mapping and eQTL colocalization to identify high-confidence risk genes, including those targeted by approved drugs or compounds in Phase II/III clinical trials. Our approach allows for the identifications of drugs that are commonly prescribed for conditions that co-exist with IBD, suggesting that these drugs could be evaluated using observational data, eliminating the need for subsequent randomized trials. These findings from our study could facilitate risk loci and genes discovery and drug development for IBD.

## RESULTS

### Identification of genetic risk loci and target genes associated with IBD

In the meta-analysis of European populations, we included a total of 49,022 cases of IBD and 1,293,489 controls (**Supplementary Table 1**) from Veterans Affairs (VA) Million Veteran Program (MVP) [35], the FinnGen study (R12)[11], Pan-UK Biobank (UKB)[36] and a GWAS meta-analysis of UK IBD Genetics Consortium (UKIBDGC) and International Inflammatory Bowel Disease Genetics Consortium (IIBDGC)[3] (Hereafter referred as IIBDGC population). Using fixed-effect meta-analysis (**Methods**), we identified 19 novel loci associated with IBD at genome-wide significance ($P < 5 \times 10^{-8}$). Subtype analyses further revealed 15 novel loci associated with CD and 27 associated with UC (**Table 1, Figure 1**).

While the genomic inflation factors ($\lambda$) showed mild inflations for summary statistics ($\lambda_{CD} = 1.08$, $\lambda_{UC} = 1.10$, and $\lambda_{IBD} = 1.12$), the linkage disequilibrium score regression (LDSC) intercepts (0.97, 0.99, and 0.99 for CD, UC, and IBD, respectively) indicated such inflation was due to polygenic effects of the variants instead of confounding population stratification. The quantile-quantile (QQ) plots of the full GWAS meta data and GWAS meta data after removing the loci that are located within 300 Kilobyte (Kb) of previously identified subtype specific GWAS significant loci were presented in **Supplementary Figure 1**. Combining evidence from LDSC and Q-Q plots showed that population structure and other confounding factors were well-controlled.

In the multi-ancestry meta-analysis, we aggregated a total of 63,415 IBD cases and 1,308,945 controls after combining the European data with an additional 14,393 cases of IBD and 15,456 controls from a recent GWAS of East Asian populations[4]. At genome-wide significance level, we identified additional 9 novel loci associated with IBD. Further stratified analysis revealed 11 novel genetic loci associated with CD, and 12 novel loci associated UC (**Table 2, Figure 2**). Combined genomic inflation factors ($\lambda = 1.07$-$1.10$), LDSC intercepts (0.98-1.00), and the Q-Q plots (**Supplementary Figure 1**) suggests that population stratification has been well controlled. The Q-Q plots for individual GWAS summary statistics were presented in **Supplementary Figure 2**.

In total, we identified 90 unique novel genetic loci associated with IBD after combining findings from the above analyses (**Supplementary Table 2**). All variants reached the genome-wide significant threshold ($P < 5\times10^{-8}$) were presented in **Supplementary Table 3** for European-specific meta-analysis and **Supplementary Table 4** for multi-ancestry meta-analysis.

To identify putative target genes in the 90 novel and the previously reported 317 GWAS-identified lead variants collected from literature (**Supplementary Table 5; Methods**), we performed xQTL mapping, including expression (eQTL), alternative splicing (AS-QTL), and alternative polyadenylation (APA-QTL) analyses, using bulk RNA-seq data from 707 European-ancestry individuals (423 from BarcUVa-Seq project[37] and 284 from GTEx[34]), and 364 Asian-ancestry individuals from our previous studies [34] (**Methods**). In addition, we also conducted cell type-specific eQTL using cell type-specific gene expression deconvoluted from bulk colon tissue gene expression[38] (**Methods**). At a Bonferroni-corrected significance threshold of $P < 0.05$, we identified 248 risk genes (corresponding 313 gene-variants pairs), including 236 genes in the European population and additional 8 genes in the East Asian population (**Supplementary Table 6**). Furthermore, colocalization analysis between GWAS and xQTL signals using *coloc*[39] identified 53 high-confidence risk genes, supported by strong evidence of shared causal signals (PP.H4 > 0.8; **Supplementary Table 7; Methods**). Of them, 13 genes are linked to lead variants from the newly identified GWAS loci.

**Identification of risk genes through multi-ancestry TWAS framework**

We next conducted multi-ancestry TWAS, AS-WAS, and APA-WAS through ancestry-specific genetic prediction model building based on the above bulk RNA-seq data (**Supplementary Table 8)**. For TWAS in European populations, we identified 318 genes significantly associated with IBD risk at Bonferroni-corrected $P < 0.05$ (**Supplementary Table 9**). Meta-analysis incorporating Asian ancestry data revealed an additional 17 significant genes. We also identified additional genes through subtype-

specific analysis. For CD, we identified 35 significant genes in the European cohort and 9 additional genes through meta-analysis (**Supplementary Table 9**). For UC, 39 genes were significant in the European cohort, with 12 additional genes identified in meta-analysis (**Supplementary Table 9**). In total, TWAS uncovered 430 IBD-associated genes, including 392 from European ancestry and 38 additional genes identified through meta-analysis with Asian ancestry. We further performed fine-mapping using FOCUS[14] to identify high-confidence genes, defined as those with a posterior inclusion probability (PIP) > 0.5 in FOCUS analyses, or, for regions not analyzed by FOCUS, as the gene with the smallest TWAS P value within a ±1 Mb window (**Methods**). This resulting in 207 high-confidence genes, with 146 from FOCUS and an additional 61 from regions beyond FOCUS analyses (**Supplementary Table 9**).

For APA-WAS, we identified 86 genes associated with IBD, including 75 in European populations and an additional 11 genes from meta-analysis of European and East Asian populations. For CD, we identified 13 additional significant genes associated with CD in the European cohort, while no significant genes were detected from the meta-analysis. For UC, APA-WAS identified 9 additional significant genes in European ancestry and 5 additional genes from meta-analysis (**Supplementary Table 10**). In total, APA-WAS identified 113 genes, including 97 from European ancestry and 16 additional genes through meta-analysis (**Supplementary Table 10**). Fine-mapping analysis prioritized 73 high-confidence genes, including 53 with PIP > 0.5 from FOCUS analyses and 20 additional genes with the smallest APA-WAS *P*-values within 1 Mb loci (**Supplementary Table 10**). The definition of high-confidence genes was presented in **Methods**.

For AS-WAS, we identified 277 genes associated with IBD, including 257 in European populations and additional 20 genes from meta-analysis of European and East Asian populations (**Supplementary Table 11**). For CD, 27 genes were significant in European ancestry, and 11 more were identified via meta-analysis. For UC, 22 genes were significant in European ancestry, and an additional 8 genes were discovered through meta-analysis. Altogether, AS-WAS identified 345 genes, comprising 309 from European ancestry and 36 from meta-analysis with Asian ancestry. Fine-mapping underscored 145 high-confidence genes, including 116 with PIP > 0.5 from FOCUS analyses and 29 additional genes with the smallest AS-WAS *P*-values within 1 Mb loci (**Supplementary Table 11).**

For cell type-specific TWAS, we first estimated the heritability contributed by SNPs for three major cell type, absorptive (ABS) cells, goblet (GOB), and stem cells (STM) using stratified LDSC analysis [40](**Methods**). The analysis revealed substantial heritability contributions from these cell types (**Supplementary Table 12).** For CD as an example, among the European ancestry, heritability was

contributed to ABS cells with 65.19% (standard error, SE: 0.06), GOB with 49.36% (SE: 0.05), and STM with 17.48% (SE: 0.02). Cell type-specific TWAS based on prediction model building in three major cell types (ABS, GOB, and STM) identified 255 genes significantly associated with IBD in European (**Supplementary Table 13**), and an additional 17 genes through meta-analysis with Asian ancestry (**Supplementary Table 13**). For the subtype CD, we identified 38 significant genes in the European cohort (**Supplementary Table 13**) and 5 additional genes via meta-analysis. For the subtype UC, we observed 34 additional significant genes in the European cohort and 7 more genes through meta-analysis. In total, the cell type-specific TWAS analysis uncovered 356 genes associated with IBD, CD and UC, including 327 from European ancestry and 29 additional genes identified through cross-ancestry meta-analysis. Fine-mapping underscored 241 high-confidence genes, including 179 with PIP > 0.5 from FOCUS analyses and 62 additional genes with the smallest cell type-specific TWAS *P*-values within 1 Mb loci.

After pooling results across the above analyses, we uncovered 802 genes that are associated with IBD, including 484 high-confidence genes. Of these, 372 genes have not been reported in previous TWAS[31,32], with 287 from TWAS, APA-WAS, and AS-WAS (**Figure 3**) and 165 from cell type-specific TWAS (**Supplementary Figure 3**). Among these, we identified 100 genes located in 43 GWAS loci that had not been previously reported (**Supplementary Table 14**). After combining the results with xQTL mapping in GWAS risk loci, we identified 506 high-confidence risk genes for IBD. Of them, 384 genes have not been reported in previous TWAS[31,32] and GWAS [3,4,9] (**Supplementary Table 15**). Of the 506 high-confidence genes, we found that analyses from these molecular traits largely provide complementary contributions to the discovery of risk genes (**Supplementary Figure 4**).

**Functional evidence of etiological roles for the newly identified risk genes**

Functional enrichment analyses on the 506 high-confidence risk genes revealed significant enrichment in the established immune and signaling pathways, such as Cytokine Signaling, Interferon Gamma Signaling and Allograft rejection pathway (**Supplementary Table 15; Methods**). In particular, these genes encompass a broad spectrum of biological processes that align with and extend beyond previously established IBD mechanisms, which included Microbial Sensing & Innate Immunity, Autophagy, IL-23/Th17 Signaling, NF-κB Signaling, Epithelial Barrier, Antigen Presentation (*HLA/ERAP*), T Cell Regulation, and Leukocyte Trafficking **(Supplementary Table 16**). These include key regulators of innate immune sensing and autophagy such as *ATG16L1* and *LRRK2*; *Th17*/*IL-23* signaling genes including *IL23R*, *STAT3*, *JAK2*, and *TYK2*; NF-κB signaling components such as *NFKB1* and *REL*; epithelial barrier and mucosal integrity genes like *MUC1* and *GNA12*; and antigen processing

and presentation molecules within the HLA region (e.g., *HLA-DQA, HLA-DRB, TAP1*, *ERAP2*). These findings not only recapitulate prior GWAS, but also integrate key functional genes implicated in IBD pathogenesis through immune modulation, microbial interaction, and epithelial homeostasis.

**Risk genes supported by scRNA-seq and spatial transcriptomics data**

We next analyzed scRNA-seq datasets on normal colon tissues from 18 non-IBD controls and inflamed colon tissues from 65 CD patients[41](**Methods**) using differential expression analysis on three major cell types (ABS, GOB, and STM). We provided evidence that the identified high-confidence genes ($P = 3.76 \times 10^{-5}$), or genes stratified by cell type-specific TWAS ($P = 3.65 \times 10^{-6}$), bulk-level TWAS approaches (TWAS, APA-WAS, and AS-WAS ($P = 4.92 \times 10^{-17}$), and were significantly enriched among the differentially expressed genes (DEGs) between non-IBD controls and CD patients (**Supplementary Table 17; Methods**). For example, the genes identified through ABS-specific TWAS were enriched among DEGs in ABS cells ($P = 3.28 \times 10^{-5}$) (**Supplementary Table 17**). Additionally, risk genes from bulk-level TWAS approaches were enriched among DEGs in all three cell types: ABS ($P = 3.15 \times 10^{-17}$), GOB ($P = 4.07 \times 10^{-8}$), and STM ($P = 6.23 \times 10^{-17}$) (**Supplementary Table 18**).

Among the 506 high-confidence genes, 268 genes (52.96%) were supported by differential expression analysis (nominal $P < 0.05$) with consistent association direction relative to TWAS or xQTL results, providing additional evidence of transcriptional dysregulation involved in IBD pathogenesis (**Supplementary Table 19**). Specifically, 122 genes (22.13%) were identified through single-cell differential expression analysis by comparing the expression between 18 non-IBD controls and 65 CD patients, while 220 genes were supported by cell type-specific expression from spatial transcriptomics data, including 146 (28.85%) genes not detected in the single-cell differential expression analysis (**Supplementary Table 19**).

**Characterizing drug target genes and therapeutic candidates**

Using data from DrugBank[42], ChEMBL[43], and the TTD[44], we comprehensively annotated our findings as therapeutic candidates targeted by approved or clinical-stage drugs. Among 506 genes, we identified 60 genes encoding druggable proteins, potentially targeted by 796 approved drugs or drugs in clinical trials (**Supplementary Table 20**). Of them, we characterized 46 genes linked to 225 drugs approved or under clinical trial phase II/III (**Figure 4**), suggesting their potential as therapeutic options for treating CD, IBD, and UC. Of note, our results found the drug Sulfasalazine already utilized to treat IBD/CD/UC and another drug Tenapanor utilized to treat irritable bowel syndrome (IBS). We highlighted these genes involved in immune-related pathways, such as the IL23-Th17-cytokine axis, as a promising

mechanism underlying IBD pathogenesis. This finding aligns with previous studies showing that IL23 supports the expansion and maintenance of Th17 cells, which secrete proinflammatory cytokines including IL17A, IL21, and IL22 that drive mucosal inflammation, neutrophil recruitment, and epithelial injury[45,46].

**DISCUSSION**

In this study, we identified 90 previously unreported loci associated with IBD through a meta-analysis of multi-ancestry GWAS data, thereby significantly expanding the known genetic architecture of IBD. To prioritize potential causal genes, we integrated large-scale, multi-ancestry, and multi-omics datasets using an innovative sTF-TWAS approach that incorporated both bulk and single-cell expression data. Fine-mapping and colocalization analyses further supported the identification of 506 high-confidence IBD risk genes, including 384 not previously reported genes. These results significantly increase the number of putative susceptibility genes and deepen our understanding of the cellular mechanisms driving IBD pathogenesis.

Beyond genes mapped to canonical immune and epithelial barrier pathways, we identified a substantial number of novel risk genes potentially involved in underexplored disease mechanisms. For example, genes involved in vesicular protein trafficking within the early secretory pathway (e.g., *TMED5*)[47] is important in maintaining Golgi apparatus structure and facilitating the transport of glycosylphosphatidylinositol-anchored proteins. Genes involved in cell adhesion and signal transduction (e.g., *ITGB3*)[48] suggest an important role in reactive oxygen species (ROS)-induced migration and invasion in colorectal cancer. *ITGB3* was also found to enrich in intestinal epithelial cells, whose elevation can induce cell senescence signatures in colonic epithelial cells[49]. Additional candidates, such as *NUCKS1*, essential in genome stability and metabolic regulation, are involved in DNA damage response and metabolism, inflammatory immune response, and microRNA[50]. The discovery revealed additional key biological processes relevant to IBD, including epithelial–immune crosstalk, stress responses, and metabolic reprogramming, providing new opportunities for functional validation and therapeutic development.

To further investigate transcriptional dysregulation in IBD, we found that more than half of the prioritized genes were supported by differential expression analysis using single-cell and spatial transcriptomic datasets from colon and IBD tissues. Several of these genes are also supported by previous experimental studies. For example, *GPX4*, a member of the glutathione peroxidase family, can detoxify lipid peroxides and regulate ferroptosis, whose inhibition can trigger ferroptotic cell death[52,53]. It can

function as antioxidants through the incorporation of selenocysteine at the catalytic site, which facilitate oxidation-reduction reactions[54]. *GSDMB* was downregulated in early enterocyte cells, its potential involvement in IBD-associated cellular dysfunction[55]. *GSDMB* expression was found to be upregulated during IBD, particularly within intestinal epithelial cells, where it is linked to gene pathways enriched for proliferation, migration, and adhesion, but not for pyroptosis[56]. *PIGR* is an epithelial glycoprotein, whose deletion may be associated with alterations in the gut microbiota and increased susceptibility to colitis[57]. *PIGR* mediates the transcytosis of secretory IgA (SIgA) across epithelial cells into the intestinal lumen, where SIgA preserves spatial segregation between the microbiota and the epithelial surface. Perturbation of the SIgA-*PIGR*-microbial axis is associated with increased risk of allergic and inflammatory diseases of the intestine[58]. *SLC44A4*'s function involves uptaking the gut microbiota-generated thiamine pyrophosphate (TPP) in the colon, so the deletion of *SLC44A4* affects colon physiology and increases the susceptibility to inflammation[59]. Future functional studies are necessary to elucidate the roles of these candidate genes in stem cell differentiation, adhesion and migration, proliferation, and metabolism/signaling, key processes implicated in IBD pathogenesis.

By analyzing drug-protein interaction databases, we identified 46 genes linked to 225 drugs that are either approved or in Phase II/III clinical trials relevant to IBD treatment. Several of the identified medications are commonly prescribed to IBD patients with comorbid conditions. For instance, Furosemide, a diuretic, is commonly prescribed for hypertension, a condition that is highly prevalent among IBD patients[60]. Similarly, Metformin is a first-line treatment for type 2 diabetes, which often co-exist with IBD. Observational studies using existing administrative claims data allow for a timelier and efficiently assessment of these drugs than randomized clinical trials. Our results provide the roadmap for future etiologic studies that can confirm or refute the utility of the identified medications. Our study findings support the therapeutic potential of these compounds, highlight opportunities for drug repurposing in IBD prevention and intervention, and contribute to advancing precision medicine-guided drug discovery for IBD.

Of note, our meta-analyses were based on summary-level statistics, with population stratification carefully controlled in the original studies, although close relatedness among few samples across cohorts cannot be fully excluded. Mild inflation was observed in the meta-analysis; however, after removing loci within 300 kb of previously reported IBD-associated variants, the Q–Q plot showed substantial improvement, indicating that the observed inflation primarily reflects polygenic effects rather than residual stratification. To bridge genetic discovery and clinical translation, we integrated comprehensive drug–protein interaction databases to prioritize druggable target genes for IBD therapy. The large number

of candidate targets highlights opportunities for drug repurposing and development. Several of the approved drugs (e.g. Furosemide, Metformin) are commonly prescribed for conditions that are highly prevalent among individuals with IBD and could be candidates for evaluation. Future studies leveraging real-world data and target trial emulation on these promising drugs, in conjunction with mechanistic experiments and clinical trials, will be critical to validate these candidates and accelerate precision-medicine–based therapies for IBD.

## METHODS

### GWAS cohorts in both European and Asian populations

In the meta-analysis of European ancestry, we obtained the GWAS summary statistics from three large biobanks: VA MVP[35], the FinnGen study (R12, November 2024)[11], Pan-UKB[36], and a published GWAS meta-analysis of IBD (IIBDGC)[3]. VA MVP is one of the largest US national cohorts containing comprehensive genotype and phenotype data from the VA electronic health record (EHR), with GWAS summary statistics released in July 2024. IBD cases from VA EHR were defined by phecodes curated by the MCP Data Core[9]. For each phecode, participants with ≥2 phecode-mapped ICD-9 or ICD-10 codes were defined as cases. Controls were defined as those with no instance of a phecode-mapped ICD-9 or ICD-10 code. The current meta-analysis included 7,837 individuals with IBD ($N_{control}$ = 438,916), 3,402 individuals with CD ($N_{control}$ = 446,631), and 5,229 individuals with UC ($N_{control}$ = 442,102) from MVP GWAS summary statistics. FinnGen is a large public-private collaboration project within Finland and international partners. It combines and analyzes large-scale genomic and health record data from Finnish biobank and national health registries. We included the GWAS summary statistics from FinnGen (R12) in our analysis, with a total of 10,960 cases of IBD ($N_{control}$ = 489,388), 2,489 cases of CD ($N_{control}$ = 497,622), and 7,220 cases of UC ($N_{control}$ = 492,160). In addition, we obtained the summary statistics of IBD from the Pan-UKB project, with a total of 5,183 cases of IBD, 2,008 cases of CD, 3,670 cases of UC, and 330,270 controls. Finally, we included a summary statistic from a GWAS meta-analysis of UKIBDGC and IIBDGC, which contained a total of 25,042 cases of IBD ($N_{control}$ = 34,915), 12,194 cases of CD ($N_{control}$ = 28,072), and 12,366 cases of UC ($N_{control}$ = 33,609). For convenience, this GWAS meta-analysis was referred to as IIBDGC in the manuscript. The diagnosis of IBD was followed by endoscopic, histopathological, and radiological criteria. In addition to the above European individuals, we included additional East Asian populations, including 14,393 cases of IBD, 7,372 cases of CD, 6,862 cases of UC, and 15,456 controls (**Supplementary Table 1**) in our multi-ancestry GWAS meta-analysis. IBD cases were diagnosed by either the European Evidence-based Consensus or the ECCO-ESGAR Guideline for Diagnostic Assessment in IBD in the East Asian cohorts.

**Quality control and association analysis in previous individual GWASs**

In MVP, the ancestry assessment was followed by genetically inferred ancestry (GIA). Variants with an imputation quality $r^2 > 0.3$ and a minor allele count (MAC) > 20 within the relevant HARE-defined ancestry group were included in the analysis[35]. Analyses were adjusted for age, sex, and 10 ancestry-specific genetic principal components (PCs). In FinnGen, association tests were performed for variants with a minimum allele count of 5 among each phenotype's cases and controls, with sex, age, 10 PCs, FinnGen chip version 1 or 2, and legacy genotyping batch included as covariates. Among the UKB, variants with INFO scores over 0.8 and an allele count of ≥ 20 were retained. Sex and age along with the first 10 PCs were included in the GWAS modelling. In the IBD Genetics Consortium GWAS meta, variants with minor allele frequency (MAF) < 0.1% and INFO < 0.4 were removed. The sites with strong evidence for heterogeneity ($I^2 > 0.9$) were discarded in the output of the fixed-effects meta-analysis. In addition, genome-wide significant variants for which the meta-analysis *P-value* was not lower than all the cohort-specific *P*-values were removed. In the East Asian GWAS, variants with an imputation quality $r^2 > 0.6$ and MAF > 0.001 were retained in the analysis, adjusting for the first ten PCs. Details about the QC and association analysis process can be found on the original studies in **Data Availability**.

**Quality control and meta-analysis of GWAS summary statistics**

We first used the LiftOver tool[61] on UCSC Genome Browser to convert GWAS summary statistics on GRCH 37 to GRCH 38. SNPs that are not able to be converted successfully were dropped from the analysis. For each GWAS summary statistics, if duplicate SNPs were found, the SNP with the smallest *P*-value was retained in the meta-analysis. In the UKB cohort, only the alternate allele frequency in cases and controls was provided, so the MAF for them was first calculated, and the MAF for the overall population was approximated by averaging the MAF among cases and controls. If the allele frequency or MAF for all populations, neither case nor controls, was not provided in the individual GWAS, we utilized 1000 Genomes Project (1KGP) on GRCh38 EUR reference panel[62] to map MAF in the datasets. For meta-analysis, we only included variants on autosomes. Genetic variants with MAF > 0.01 remained for downstream analysis.

We conducted an inverse variance weighted meta-analysis of existing GWAS summary statistics in European populations, including data from MVP, FinnGen (R12), Pan-UKB, and IBDGC GWAS meta-analysis of IBD using METAL (release 2011-03-25)[63]. Cochran's Q Statistic and $I^2$ statistics were also calculated using METAL. In the multi-ancestry meta-analysis, we further combined data from East Asians using the same model. Genomic control was applied to each individual GWAS in METAL and

then applied to the GWAS meta-analyzed summary statistics according to the METAL document from Code Availability.

Genomic inflation factors and Q-Q plot were utilized to evaluate the genomic inflation of GWAS. If $\lambda$ close to 1, it suggests that genomic inflations were well-controlled. We next applied LDSC regression[64] by regressing the $\chi^2$ statistics of each SNP from GWAS summary statistics on its corresponding LD score. The pre-computed LD scores were obtained from the 1KGP reference panel. LDSC intercepts were also considered when evaluating genomic inflation. If the intercepts from LDSC regression show values close to 1, then the deviation of Q-Q plot may be due to polygenic effects instead of population stratification. To evaluate such effects, we further plotted the Q-Q plot by removing the loci that are located within 300 Kb of previously identified GWAS loci. If the Q-Q plot of removing previously identified loci shows minimal deviation, then the confounding of population stratification was well-controlled.

**Previous GWAS loci collection and novel locus definition**

The novel loci were defined following the previous GWASs of IBD[3,4]. We first created a reference risk SNP list for each trait (CD, UC, or IBD) from previous individual GWASs or IBD consortia that were included in the current meta-analysis[3,4,11,35,36] by selecting the SNPs that reached the genome-wide significant threshold ($P < 5\times10^{-8}$). If an SNP was reported in multiple GWASs, the one with the lower *P*-value was retained. For the reference SNP list of IBD, variants previously reported as CD- or UC- associated were also included in the list and considered as IBD-associated, while a known CD-associated SNP was not considered as associated with UC, or vice versa[3]. Next, we selected the genome-wide significant ($P < 5\times10^{-8}$) SNPs from our meta-analysis and removed the variants on the reference SNP list. To ensure the putatively identified genome-wide significant SNPs in our meta-analysis were not due to LD with previously reported variants, we further checked whether they are at least 500 Kb away from already known SNPs. If yes, the SNP was not considered novel and was removed. For the remaining genome-wide significant SNPs, we performed a LD clumping with an LD window of 500 Kb and $r^2$ threshold of 0.6 using the 1KGP EUR reference panel for European-specific meta-analysis to identify the novel genetic loci. If the genetic loci after LD clumping overlapped with other loci within 500 Kb, we then merged the loci and kept the variants with the strongest association (the lowest *P*) as the lead variants for the merged genetic loci. Clumping was performed using PLINK 1.9[65]. Finally, the retained SNPs were further cross-checked with GWAS loci identified from the latest GWASs of IBD[4] to confirm that they are novel.

For the multi-ancestry meta-analysis, the 1KGP European reference panel and 1KGP East Asian reference panel were used for LD clumping, respectively. LD clumping results from the European and East Asian panel were merged, and the lead variants with the lowest *P* within 500 Kb were kept. To identify additional novel loci from European-specific meta-analysis, we merged the overlapping loci from European-specific and multi-ancestry meta-analysis within 500 Kb and then compared the *P*-values for the lead variants in the overlapping loci. Loci where lead variants from multi-ancestry meta-analysis with the lower *P*-values were considered as the additional novel loci from multi-ancestry meta-analysis.

**RNA-seq data processing and gene expression, AS and APA quantification**

We developed ancestry-specific TWAS models using three bulk gene expression datasets: 707 European-ancestry samples, including 423 from the BarcUVa-Seq project[37] and 284 from the GTEx project[37,66], along with 364 Asian-ancestry samples from our previous study[34]. Gene-level expression was quantified using RNA-SeQC, aligning with the annotation used in GTEx as described in our previous study[37]. In general, we applied quantile and inverse normalization to the RNA-seq transcript per million (TPM) data. Next, we adjusted potential confounding factors through a probabilistic estimation of expression residuals (PEER) analysis, following the GTEx recommendations for determining the number of PEER factors. We quantified AS level from RNA-seq data using LeafCutter[67] and RegTools[67], following the methods descried in our previous study[37]. The level of APA was assessed using the percentage of PDUI, as estimated with DaPars v2.0[68], in accordance with our prior research[19].

For downstream gene expression, AS and APA prediction model building, we adjusted these measurements for potential confounding factors, including PEER factors (60 for BarcUVa-Seq and Asian-ancestry data, 45 for GTEx), as well as age, sex, and the first five genotype PCs.

**Cell type-specific gene expression quantification (deconvolution bulk RNA-seq based on the reference single-cell data.)**

We obtained scRNA-seq data for 33,806 epithelial cells from normal colon tissue using the COLON MAP project[33]. These epithelial cells are clustered into seven clusters: absorptive cells (ABS), crypt top cells (CT), enteroendocrine cells (EE), goblet cells (GOB), stem cells (STM), transit amplifying cells (TAC), and tuft cells (TUF). We applied quality control to select high quality cells: nUMI > 500, nGene > 250, log10GenesPerUMI > 0.8, and mitoRatio < 0.2. Following the CIBERSORTx[27] pipeline, we constructed a signature gene expression matrix for each cell type and estimated each cell type fraction for bulk RNA-seq data from 707 European and 364 East Asian individuals. Among these analyzed cell types, absorptive (ABS) cells exhibited the highest median proportion (49%), followed by stem-like

(STM) cells (29%), goblet (GOB) cells (3%), and other cell types (< 1%). We kept the top three most abundant cell types (ABS, STM, and GOB) and performed deconvolution of bulk RNA-seq data using their respective signature matrices to derive cell type-specific gene expression profiles. The downstream processing and normalization of cell type-specific gene expression data were consistent with those used for bulk gene expression prediction. Model building for cell type-specific expression also incorporated the recommended PEER factors, along with age, sex, and the first five genotype PCs.

**Genotype data processing**

The process of genotype data from BarcUVa-Seq and GTEx project was described in our previous study[37]. In general, 400,000 SNPs from the BarcUVa-Seq project was imputed using TOPMed (Version 2) with European ancestry as the reference panel. We then excluded variants with an imputation quality below $R^2 < 0.3$, missingness rate greater than 10%, a MAF less than 0.01 (MAF < 0.01), or significant deviations from Hardy-Weinberg equilibrium ($P < 10^{-6}$). The same process was applied to genotype data from East Asian data, which was imputed using TOPMed with Asian ancestry as reference panel[69]. In addition, GTEx project provides the whole genome sequencing (WGS) data and we downloaded the VCF format from dbGaP (phs000424.v8.p2). A comprehensive outline of WGS variant quality control procedures is available from the GTEx consortium[70].

**Identifying putative target genes for GWAS risk loci through xQTL and colocalization analyses**

To identify putative target genes associated with both novel and previously established GWAS-identified risk variants, we performed xQTL mapping using normalized bulk gene expression, AS, APA, and cell type-specific gene expression data. Linear regression analyses were conducted between each risk variant and molecular traits (eQTL, aQTL, apaQTL, or cell type-specific eQTL) for genes located within ±1 Mb of the variant, adjusting for potential confounders as described in the preceding section. For cell type-specific eQTL analyses, we focused on expression data from ABS, GOB, and STM cells. Since the majority of GWAS risk variants were identified in European populations, the corresponding xQTL analyses were primarily conducted using molecular data from individuals of European ancestry. For additional risk variants identified through meta-analysis of European and East Asian populations, we also performed xQTL analyses using data from East Asian individuals.

Statistical significance for xQTLs was determined using a Bonferroni-corrected $P < 0.05$. For these significant xQTLs, we conducted colocalization analyses to evaluate the likelihood of shared causal variants between each molecular trait and the GWAS signal. Within +/-500 Kb of each GWAS risk variant, we extracted GWAS summary statistics and corresponding xQTL data, analyzing only regions

with at least 50 overlapping variants. We employed default priors and utilized the 'coloc.abf' function to calculate the posterior probability (PP) of colocalization. Posterior probability for hypothesis 4 (PP.H4), which reflects a shared causal variant between the two traits, was used to assess colocalization. A PP.H4 > 0.8 was considered evidence of colocalization. ABF colocalization was conducted using the package "coloc" in R[39].

**TWAS, AS-WAS, APA-WAS and cell type-specific TWAS**

***Genetic prediction models***

We applied our sTF-TWAS framework[17,18] to train prediction models for gene expression, AS, and APA levels. This approach utilizes cis-genetic variants located within a ±1 Megabase (Mb) region and occupied by transcription factors[34] and the elastic-net is used as the predictive model:

$$E_g \sim \sum_{i}^{p} w_{i,g} X_i + \varepsilon$$

where $E_g$ represents the $g$-th gene expression (AS, APA or cell type-specific gene expression) level; $X_i$ denotes the genotype vector of the $i$-th variant ($X_i$=0, 1, or 2) and $p$ is the total count of variants within the cis-region. The term $w_{i,g}$ represents the effect size of $i$-*th* variant for $g$-*th* gene expression, and $\varepsilon$ is the residual with variance $\sigma_e^2$. An elastic-net penalty, combining $L_1$and $L_2$ regularizers, is applied to optimize ($w_{i,g}$) for genetic variants.

To optimize model parameters and evaluate predictive efficacy, we used five-fold cross-validation. Model performance was assessed by calculating the squared correlation ($R^2$) between predicted and measured gene expression (or AS, APA or cell type-specific gene expression).

***Association analysis***

Under our sTF-TWAS framework, we applied our ancestry-specific prediction models (gene, expression, AS, APA and cell type-specific gene expression) into GWAS summary statistics from European (49,022 IBD cases and 1,293,489 controls) and East Asian populations (14,393 cases and 15,456 controls), respectively, using the

$$\mathrm{Z}_g \approx \sum_{i \in \mathrm{Model}_g} \widehat{w}_{i,g} \frac{\hat{\sigma}_i}{\hat{\sigma}_g} \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}$$

where, the $\mathrm{Z}_g$ estimates the association between genetic prediction expression (or AS, APA, cell type-specific gene expression) and IBD risk; $\widehat{w}_{i,g}$ is the weight of $i$-th genetic variant for predicting $g$-th gene

expression. $\hat{\beta}_i$ and se ($\hat{\beta}_i$) are the GWAS-reported regression coefficients and its SE for $i$-th variant; and $\hat{\sigma}_i$ and $\hat{\sigma}_g$ are the estimated variances of $i$-th variant and the predicted expression of $g$-th gene ( or AS, APA, cell type-specific gene expression).

We conducted TWAS, AS-WAS, APA-WAS, and cell type-specific TWAS separately in European and East Asian populations. To integrate gene-level association signals across ancestries and enhance statistical power, we performed meta-analysis using the Aggregated Cauchy Association Test (ACAT). Parallel analyses were conducted for CD and UC, respectively. Statistical significance for TWAS, AS-WAS, APA-WAS, or cell type-specific TWAS associations was determined using a Bonferroni-corrected P-value threshold of < 0.05.

***Fine-mapping analysis***

We performed fine-mapping analysis to nominate putative causal genes using FOCUS[14], for TWAS, AS-WAS, APA-WAS, or cell-type specific TWAS, respectively. The primary association signal for each gene in TWAS (or other molecular trait) was first identified based on ancestry-specific results from either the European or East Asian population. Linkage disequilibrium (LD) reference panels from the corresponding populations in the 1KGP[62] were used. Fine-mapping was then conducted separately for each ancestry-driven gene set using FOCUS-defined genomic regions and ancestry-matched LD panels. Each gene was assigned a posterior inclusion probability (PIP), representing the likelihood of being causal for the observed TWAS signal.

**Stratified LDSC regression analysis**

To identify the most disease-relevant cell types in colon tissue for IBD and its subtypes, we performed a stratified LDSC by using the single-cell normalized gene expression matrix and GWAS summary statistics of IBD. The scRNA-seq gene expression data on terminal ileum (TI) and ascending colon (AC) tissues were obtained from 202,359 cells from 170 specimens across non-IBD controls (N = 18) and active/inactive CD patients (N = 65) from the National Center for Biotechnology Information Gene Expression Omnibus (GEO)[41]. The cells were classified into 16 cell types within the epithelial compartment: early enterocytes, intermediate enterocyte, mature enterocyte, early colonocyte, intermediate colonocytes cells, mature colonocyte, early GOB, mature GOB, GOB proliferating cells, cycling TA cells, BEST4/OTOP2 cells, tuft cells, enteroendocrine cells (EEC), Paneth cells, STM cells, and LND cells (annotated by high expression of *LCN2*, *NOS2*, and *DUOX2*)[41]. Among them, enterocytes and colonocytes were merged as ABS cells. ABS, GOB, and STM were considered as major cell types for IBD and included in the follow-up analysis. Using Seurat v5, we generated a pseudo bulk gene expression matrix for each of the 83 subjects (18 non-IBD controls and 65 CD patients). Pseudo bulk aggregation

was used to sum raw counts across all cells of a given type within each subject, reducing the influence of outlier cells and producing more robust, cell type-specific expression profiles[33]. Genes were then ranked by t-statistic within each cell type, and the top 10% were defined as specifically expressed genes[40]. Finally, we applied stratified LDSC on specifically expressed genes to estimate partitioned heritability and identify cell types significantly enriched for IBD risk.

**Pathway enrichment analysis and function categorization**

To explore the biological functions and pathways enriched in the gene list, we performed gene set enrichment analysis using the Enrichr web-based tool (https://maayanlab.cloud/Enrichr/). Enrichr integrates multiple curated databases[71-73], enabling comprehensive functional annotation and pathway enrichment analysis. Enrichr uses a combined score metric that incorporates the *P*-value from the Fisher exact test and the z-score of the deviation from the expected rank to rank enriched terms. Significantly enriched pathways were identified based on FDR-corrected $P < 0.05$. To contextualize these genes within known biological processes, we mapped the identified genes to established IBD-related pathways. Pathway definitions were obtained from curated databases, including KEGG, Reactome, and the Immunology Database and Analysis Portal (ImmPort). Additionally, we referenced literature-supported functional experiments that validate gene involvement in IBD pathogenesis. We further annotated genes with supporting evidence from functional studies, including expression profiling, knockout or knockdown models, and immunological assays, to highlight genes with demonstrated mechanistic roles in IBD.

**Differential expression analysis using single-cell RNA-sequencing and spatial transcriptomics**

Pseudo-bulk aggregation with summing raw counts for each cell type from 18 non-IBD controls and 65 CD patients was used, as described in the preceding section. We performed differential expression analysis for each cell type between non-IBD controls and active/inactive CD subjects using DESeq2 with default parameters. In addition, we compared the gene expression of each individual cell type to the rest of other cells within both non-IBD controls and active/inactive CD patients. Genes with FDR-corrected $P < 0.05$ were considered as significantly differentially expressed genes. If a significant gene identified by TWAS and xQTL also showed an average log2 fold change in the same direction in the single-cell analysis, and had a nominal *P*-value < 0.05, it was considered to have a consistent direction of association with the TWAS and xQTL results.

Differential expression analysis was also performed in spatial transcriptomic data obtained from CD samples from 10X Genomics. More detailed information for spatial transcriptomic data processing can be found in a previous study[41]. Four CD samples from the above samples were selected because of

their relatively high proportions of LND cells[41]. Among them, two samples were with active TI disease (GCA092 and GCA033), while the other two were with active AC disease (GCA089 and GCA099)[41]. Data normalization and scaling were performed in R using SCTransform and ScaleData functions, respectively. To identify spatially variable genes, we used FindVariableFeatures function to select the most informative genes for downstream analysis. Principal component analysis (PCA) was used to reduce dimensionality using RunPCA function, and the top 10 PCs were used for clustering with FindClusters function at a resolution of 0.8. FindNeighbors function was used to compute a nearest-neighbor graph based on PCA. For differential expression analysis, the FindAllMarkers function was used to compare the gene expression profile between one cluster and all others within the same sample, with the minimum percentage expression threshold of 0.25 and log fold-change threshold of 0.25. *P*-values were adjusted using Bonferroni correction, and corrected $P < 0.05$ were considered as statistically significant genes. Uniform Manifold Approximation and Projection (UMAP) were performed for visualization.

Cell type annotation for each cluster was conducted using GPTCelltype[74] by selecting the top positively expressed feature genes. Clustering visualization was presented in **Supplementary Figure 5**. Combing results from four samples, 3,047 unique genes (48.26%) were differentially expressed (Bonferroni-corrected $P < 0.05$) among the total 7,060 genes.

**Characterizing potential therapeutic drug candidates**

To identify potential therapeutic drug candidates, we assembled a comprehensive set of drug-target pairs from four drug databases: DrugBank[42], ChEMBL[43], and the TTD[44]. DrugBank is a curated database containing over 500,000 drugs and thousands of bioentities (proteins, genes, and other cellular components). ChEMBL provides detailed information on more than 2 million bioactivity assays for bioactive molecules with drug-like properties, covering over 2 million compounds. TTD offers comprehensive information on thousands of drug targets, including 426 approved, 1,014 in clinical trials, 212 in preclinical or patented stages, and 1,479 reported in the literature. Drugs and their paired genes were comprehensively downloaded from these three databases. We subsequently linked these drugs to our identified disease-susceptibility genes.

## DATA AVAILABILITY

All GWAS summary statistics included in the current study are publicly available. VA MVP summary statistics is available at dbGAP (accession number phs002453). FinnGen summary statistics can be accessed at https://finngen.gitbook.io/documentation/data-download. The Pan-UK Biobank GWAS summary statistics is available at https://pan.ukbb.broadinstitute.org. IBDGC European GWAS meta-

analysis can be downloaded through ftp://ftp.sanger.ac.uk/pub/project/humgen/summary_statistics/human/2016-11-07/. The East Asian GWAS summary statistics is available at IIBDGC website: https://www.ibdgenetics.org. ScRNA-seq datasets are available at the GEO database under accession number GSE266546. Spatial transcriptomic data are available at request to corresponding authors.

LDSC Reference data is available at https://alkesgroup.broadinstitute.org/LDSCORE.

## CODE AVAILABILITY

The code used our study is available at GitHub: https://github.com/XingyiGuo/IBD_genetics.

PLINK 1.9: Shaun Purcell, Christopher Chang; www.cog-genomics.org/plink/1.9/

METAL: GitHub - statgen/METAL: Meta-analysis of genomewide association scans

METAL Documentation: https://genome.sph.umich.edu/wiki/METAL_Documentation

LDSC v1.0.1: https://github.com/bulik/ldsc

Seurat: https://satijalab.org/seurat/

Enrichr web-based tool: https://maayanlab.cloud/Enrichr/

## ACKNOWLEDGEMENTS

We acknowledge the significant genetic resource contributions from previous IBD GWASs, VA MVP, FinnGen study, UKB, UKIBDGC and IIBDGC.

## FUNDING SOURCE

This research was supported by US National Institutes of Health (NIH) grant R01CA269589-01A1 to X.G. and R01DK103831 to K, L.

## REFERENCES

1. Ng, S.C. *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* **390**, 2769-2778 (2018).
2. Lewis, J.D. *et al.* Incidence, Prevalence, and Racial and Ethnic Distribution of Inflammatory Bowel Disease in the United States. *Gastroenterology* **165**, 1197-1205 e2 (2023).
3. de Lange, K.M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* **49**, 256-261 (2017).
4. Liu, Z. *et al.* Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat Genet* **55**, 796-806 (2023).
5. Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182 (2024).
6. Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* **40**, 955-62 (2008).
7. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-24 (2012).
8. Levine, A.P. *et al.* Genetic Complexity of Crohn's Disease in Two Large Ashkenazi Jewish Families. *Gastroenterology* **151**, 698-709 (2016).

9. Liu, J.Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-986 (2015).
10. Liu, T.C. & Stappenbeck, T.S. Genetics and Pathogenesis of Inflammatory Bowel Disease. *Annu Rev Pathol* **11**, 127-48 (2016).
11. Kurki, M.I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508-518 (2023).
12. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091-8 (2015).
13. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245-52 (2016).
14. Mancuso, N. *et al.* Probabilistic fine-mapping of transcriptome-wide association studies. *Nat Genet* **51**, 675-682 (2019).
15. Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* **51**, 592-599 (2019).
16. Amariuta, T. *et al.* IMPACT: Genomic Annotation of Cell-State-Specific Regulatory Elements Inferred from the Epigenome of Bound Transcription Factors. *Am J Hum Genet* **104**, 879-895 (2019).
17. Wen, W. *et al.* Genetic variations of DNA bindings of FOXA1 and co-factors in breast cancer susceptibility. *Nat Commun* **12**, 5318 (2021).
18. He, J. *et al.* Integrating transcription factor occupancy with transcriptome-wide association analysis identifies susceptibility genes in human cancers. *Nat Commun* **13**, 7118 (2022).
19. Guo, X. *et al.* Large-Scale Alternative Polyadenylation-Wide Association Studies to Identify Putative Cancer Susceptibility Genes. *Cancer Res* **84**, 2707-2719 (2024).
20. Goldman, S.L. *et al.* The Impact of Heterogeneity on Single-Cell Sequencing. *Front Genet* **10**, 8 (2019).
21. Aygun, N. *et al.* Brain-trait-associated variants impact cell-type-specific gene regulation during neurogenesis. *Am J Hum Genet* **108**, 1647-1668 (2021).
22. Song, X. *et al.* MiXcan: a framework for cell-type-aware transcriptome-wide association studies with an application to breast cancer. *Nat Commun* **14**, 377 (2023).
23. Thompson, M. *et al.* Multi-context genetic modeling of transcriptional regulation resolves novel disease loci. *Nat Commun* **13**, 5704 (2022).
24. Li, Q. *et al.* Molecular profiling of human non-small cell lung cancer by single-cell RNA-seq. *Genome Med* **14**, 87 (2022).
25. Tian, Y. *et al.* Single-cell transcriptomic profiling reveals the tumor heterogeneity of small-cell lung cancer. *Signal Transduct Target Ther* **7**, 346 (2022).
26. Wu, F. *et al.* Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat Commun* **12**, 2540 (2021).
27. Newman, A.M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773-782 (2019).
28. Bryois, J. *et al.* Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat Neurosci* **25**, 1104-1112 (2022).
29. Yazar, S. *et al.* Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**, eabf3041 (2022).
30. Perez, R.K. *et al.* Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science* **376**, eabf1970 (2022).
31. Mancuso, N. *et al.* Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *Am J Hum Genet* **100**, 473-487 (2017).
32. Diez-Obrero, V. *et al.* Transcriptome-Wide Association Study for Inflammatory Bowel Disease Reveals Novel Candidate Susceptibility Genes in Specific Colon Subsites and Tissue Categories. *J Crohns Colitis* **16**, 275-285 (2022).

33. Chen, B. *et al.* Differential pre-malignant programs and microenvironment chart distinct paths to malignancy in human colorectal polyps. *Cell* **184**, 6262-6280 e26 (2021).
34. Zhishan Chen, W.S., Qing Li, et al. Identifying putative susceptible transcription factors and genes for colorectal cancer: an integrated analysis of large-scale genetic and multi-omics data. (2025).
35. Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182 (2024).
36. Karczewski, K.J. *et al.* Pan-UK Biobank GWAS improves discovery, analysis of genetic architecture, and resolution into ancestry-enriched effects. *medRxiv*, 2024.03.13.24303864 (2024).
37. Chen, Z. *et al.* Novel insights into genetic susceptibility for colorectal cancer from transcriptome-wide association and functional investigation. *J Natl Cancer Inst* **116**, 127-137 (2024).
38. Qing Li, J.W., Victor Moreno, Chao Li, Jungyoon Choi, Zhishan Chen, Wanqing Wen, Xiang Shu, Xiao-ou Shu, Jirong Long, Qiuyin Cai, Bhuminder Singh, Ken S Lau, Ulrike Peters, Graham Casey, Wei Zheng, Xingyi Guo. Multi-ancestry transcriptome-wide association analyses on single-cell deconvoluted expression data to identify cell-type specific risk genes for colorectal cancer. in *American Society of Human Genetics 2024* (Denver, Colorado, 2024).
39. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
40. Finucane, H.K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629 (2018).
41. Li, J. *et al.* Identification and multimodal characterization of a specialized epithelial cell type associated with Crohn's disease. *Nat Commun* **15**, 7204 (2024).
42. Knox, C. *et al.* DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Res* **52**, D1265-D1275 (2024).
43. Zdrazil, B. *et al.* The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res* **52**, D1180-D1192 (2024).
44. Zhou, Y. *et al.* TTD: Therapeutic Target Database describing target druggability information. *Nucleic Acids Res* **52**, D1465-D1477 (2024).
45. Abraham, C. & Cho, J. Interleukin-23/Th17 pathways and inflammatory bowel disease. *Inflamm Bowel Dis* **15**, 1090-100 (2009).
46. Kobayashi, T. *et al.* IL23 differentially regulates the Th1/Th17 balance in ulcerative colitis and Crohn's disease. *Gut* **57**, 1682-9 (2008).
47. Zhou, L. *et al.* TMED family genes and their roles in human diseases. *Int J Med Sci* **20**, 1732-1743 (2023).
48. Lei, Y. *et al.* Proteomics identification of ITGB3 as a key regulator in reactive oxygen species-induced migration and invasion of colorectal cancer cells. *Mol Cell Proteomics* **10**, M110 005397 (2011).
49. Tun, X. *et al.* Integrin beta3-Mediated Cell Senescence Associates with Gut Inflammation and Intestinal Degeneration in Models of Alzheimer's Disease. *Int J Mol Sci* **24**(2023).
50. Huang, P., Cai, Y., Zhao, B. & Cui, L. Roles of NUCKS1 in Diseases: Susceptibility, Potential Biomarker, and Regulatory Mechanisms. *Biomed Res Int* **2018**, 7969068 (2018).
51. Qiu, Z., Khairallah, C., Romanov, G. & Sheridan, B.S. Cutting Edge: Batf3 Expression by CD8 T Cells Critically Regulates the Development of Memory Populations. *J Immunol* **205**, 901-906 (2020).
52. Xu, S. *et al.* The emerging role of ferroptosis in intestinal disease. *Cell Death Dis* **12**, 289 (2021).
53. Mayr, L. *et al.* Dietary lipids fuel GPX4-restricted enteritis resembling Crohn's disease. *Nat Commun* **11**, 1775 (2020).
54. Short, S.P. *et al.* Colonic Epithelial-Derived Selenoprotein P Is the Source for Antioxidant-Mediated Protection in Colitis-Associated Cancer. *Gastroenterology* **160**, 1694-1708 e3 (2021).

55. Privitera, G., Rana, N., Armuzzi, A. & Pizarro, T.T. The gasdermin protein family: emerging roles in gastrointestinal health and disease. *Nat Rev Gastroenterol Hepatol* **20**, 366-387 (2023).
56. Rana, N. *et al.* GSDMB is increased in IBD and regulates epithelial restitution/repair independent of pyroptosis. *Cell* **185**, 283-298 e17 (2022).
57. Bruno, M.E. *et al.* Correlation of Biomarker Expression in Colonic Mucosa with Disease Phenotype in Crohn's Disease and Ulcerative Colitis. *Dig Dis Sci* **60**, 2976-84 (2015).
58. Kaetzel, C.S. Cooperativity among secretory IgA, the polymeric immunoglobulin receptor, and the gut microbiota promotes host-microbial mutualism. *Immunol Lett* **162**, 10-21 (2014).
59. Sabui, S. *et al.* Effect of knocking out mouse Slc44a4 on colonic uptake of the microbiota-generated thiamine pyrophosphate and colon physiology. *Am J Physiol Gastrointest Liver Physiol* **327**, G36-G46 (2024).
60. Seif, F. *et al.* The role of JAK-STAT signaling pathway and its regulators in the fate of T helper cells. *Cell Commun Signal* **15**, 23 (2017).
61. Hinrichs, A.S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**, D590-8 (2006).
62. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
63. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
64. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
65. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
66. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
67. Cotto, K.C. *et al.* Integrated analysis of genomic and transcriptomic data for the discovery of splice-associated variants in cancer. *Nat Commun* **14**, 1589 (2023).
68. Li, L. *et al.* An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat Genet* **53**, 994-1005 (2021).
69. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021).
70. Consortium, G.T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
71. Chen, E.Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
72. Kuleshov, M.V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-7 (2016).
73. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**, e90 (2021).
74. Hou, W. & Ji, Z. Assessing GPT-4 for cell type annotation in single-cell RNA-seq analysis. *Nat Methods* **21**, 1462-1465 (2024).

**Figure legends**

**Figure 1. Manhattan plot of GWAS meta-analysis results in the European populations.** The plot shows the novel loci identified from a meta-analysis of 49,022 IBD cases and 1,293,489 controls of European ancestry.

**Figure 2. Manhattan plot of GWAS meta-analysis results across European and East Asian populations.** The plot shows the novel loci identified from a meta-analysis including 49,022 IBD cases and 1,293,489 controls of European ancestry, along with 14,393 IBD cases and 15,456 controls of East Asian ancestry.

**Figure 3. Novel risk genes identified across IBD, CD, and UC from TWAS, APA-WAS, and AS-WAS analyses.** A total of 287 novel risk genes were discovered and categorized by disease specificity: "IBD" (unique to IBD), "CD" (unique to Crohn's disease), "UC" (unique to ulcerative colitis), "CD,IBD" (shared by CD and IBD), "CD,UC" (shared by CD and UC), "IBD,UC" (shared by IBD and UC), and "CD,IBD,UC" (shared by all three). Gene names are truncated with remaining characters replaced by an asterisk (*) due to space constraints.

**Figure 4. Circular plot of druggable genes and associated drugs for IBD treatment.**
The plot displays 95 druggable proteins potentially targeted by 388 approved drugs or compounds currently in clinical trials for IBD. From the inner to outer layers, the plot shows IBD, proteins, and their corresponding drugs.

**Supplementary Figures**

**Supplementary Figure 1. Quantile-quantile (QQ) plots of GWAS meta-analysis results.**
QQ plots are shown for the full GWAS meta-analysis dataset and for the dataset after excluding loci within 300 kilobases (Kb) of previously reported genome-wide significant loci for each IBD subtype.

**Supplementary Figure 2. The Q-Q plots of individual GWAS summary statistics included in the meta-analysis.**
QQ plots are shown for individual GWAS summary statistics included in the meta-analysis for each IBD subtype.

**Supplementary Figure 3. Cell type-specific TWAS identifies 165 novel risk genes across IBD, CD, and UC.**
A total of 165 novel risk genes were discovered and categorized by disease specificity: "IBD" (unique to IBD), "CD" (unique to Crohn's disease), "UC" (unique to ulcerative colitis), "CD,IBD" (shared by CD and IBD), "CD,UC" (shared by CD and UC), "IBD,UC" (shared by IBD and UC), and "CD,IBD,UC" (shared by all three). Gene names are truncated with remaining characters replaced by an asterisk (*) due to space constraints.

**Supplementary Figure 4. Unique and shared high-confidence genes identified from molecular analyses for IBD, CD, UC, and their subtypes.**

The horizontal bars represent the total number of genes identified by each method. A solid circle denotes genes uniquely identified by that method. Circles connected by a black line indicate genes shared exclusively between those methods but not identified by others.

**Supplementary Figure 5**. **Clustering visualization of spatial transcriptomic data.**
Cell type annotation produced by using GPTCelltype by selecting the top positively expressed feature genes.

**Tables**

**Table 1. Novel genetic risk variants identified in the European GWAS meta-analysis.**

| Disease | Chr[a] | Position | Lead variant | REF[b] | Allele1[c] | Allele2[d] | Effect | SE[e] | *P* | MAF[f] |
|---|---|---|---|---|---|---|---|---|---|---|
| CD | 1 | 119880078 | rs79332994 | G | A | G | 0.239 | 0.039 | 8.89E-10 | 0.016 |
| CD | 2 | 111153117 | rs6734942 | C | T | C | -0.074 | 0.013 | 8.43E-09 | 0.347 |
| CD | 2 | 202623726 | rs10189685 | G | A | G | 0.075 | 0.013 | 1.39E-08 | 0.301 |
| CD | 5 | 172897975 | rs564349* | G | A | G | -0.075 | 0.013 | 1.17E-08 | 0.294 |
| CD | 6 | 42031620 | rs13211412 | A | A | T | -0.109 | 0.015 | 1.27E-12 | 0.217 |
| CD | 8 | 42493658 | rs6988165 | C | C | G | -0.068 | 0.012 | 3.89E-08 | 0.428 |
| CD | 14 | 103487329 | rs55859054 | T | A | T | -0.071 | 0.013 | 3.40E-08 | 0.358 |
| CD | 15 | 77532779 | rs16968895 | A | A | G | 0.070 | 0.013 | 2.70E-08 | 0.401 |
| CD | 15 | 90641566 | rs7171533 | G | A | G | -0.115 | 0.020 | 9.19E-09 | 0.098 |
| CD | 17 | 7104303 | rs140313246 | G | A | G | -0.169 | 0.028 | 1.45E-09 | 0.093 |
| CD | 17 | 40558934 | rs2228015 | T | T | C | -0.199 | 0.036 | 4.40E-08 | 0.050 |
| CD | 18 | 48868651 | rs7240004* | A | A | G | 0.075 | 0.013 | 4.23E-09 | 0.374 |
| CD | 19 | 2477822 | rs7354 | C | A | C | 0.080 | 0.014 | 6.24E-09 | 0.370 |
| CD | 19 | 40706707 | rs2115239 | G | A | G | -0.076 | 0.012 | 5.77E-10 | 0.448 |
| CD | 21 | 39094373 | rs2836881 | G | T | G | -0.099 | 0.014 | 2.31E-12 | 0.258 |
| UC | 1 | 119987477 | rs12132511 | A | A | T | -0.182 | 0.029 | 6.00E-10 | 0.017 |
| UC | 1 | 155192276 | rs4072037 | C | T | C | 0.059 | 0.010 | 3.44E-09 | 0.452 |
| UC | 1 | 169758741 | rs4265482 | T | T | C | -0.065 | 0.011 | 5.58E-09 | 0.284 |
| UC | 1 | 197732862 | rs16841904 | C | T | C | 0.069 | 0.012 | 2.29E-08 | 0.188 |
| UC | 2 | 28404762 | rs11127150 | C | T | C | -0.061 | 0.010 | 8.18E-10 | 0.425 |
| UC | 2 | 43388869 | rs6739828 | T | T | G | -0.058 | 0.010 | 3.89E-09 | 0.443 |
| UC | 4 | 102690688 | rs227368 | C | T | C | -0.056 | 0.010 | 9.20E-09 | 0.497 |
| UC | 4 | 155679826 | rs4691843 | T | A | T | 0.058 | 0.010 | 1.03E-08 | 0.438 |
| UC | 5 | 131087866 | rs4705885 | G | A | G | 0.069 | 0.011 | 2.13E-10 | 0.314 |
| UC | 5 | 131726714 | rs12523195 | T | T | C | -0.066 | 0.010 | 2.07E-10 | 0.334 |
| UC | 6 | 43831244 | rs57266556 | C | T | C | 0.080 | 0.014 | 8.37E-09 | 0.161 |
| UC | 6 | 90268442 | rs62408234 | C | T | C | -0.088 | 0.014 | 6.87E-10 | 0.183 |
| UC | 7 | 20542073 | rs76380568 | T | T | C | 0.109 | 0.019 | 9.43E-09 | 0.075 |
| UC | 8 | 58856561 | rs2726552 | G | A | G | -0.062 | 0.010 | 4.38E-10 | 0.464 |
| UC | 8 | 80353593 | rs1911715 | G | A | G | 0.064 | 0.010 | 5.93E-10 | 0.355 |
| UC | 8 | 128060586 | rs2608056 | A | A | G | 0.063 | 0.011 | 2.96E-09 | 0.303 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| UC | 11 | 78560568 | rs6592792 | A | A | C | 0.074 | 0.012 | 2.79E-09 | 0.189 |
| UC | 13 | 99348586 | rs9585037 | A | A | T | 0.077 | 0.013 | 1.27E-09 | 0.222 |
| UC | 16 | 67646903 | rs117556162 | G | A | G | 0.123 | 0.022 | 1.87E-08 | 0.038 |
| UC | 16 | 80607068 | rs7195620 | G | A | G | 0.062 | 0.011 | 2.26E-08 | 0.310 |
| UC | 17 | 47441217 | rs62076510 | T | T | G | -0.077 | 0.014 | 1.38E-08 | 0.158 |
| UC | 18 | 49646539 | rs72915967 | T | T | C | -0.065 | 0.011 | 9.28E-09 | 0.239 |
| UC | 19 | 43648948 | rs4760 | A | A | G | -0.077 | 0.014 | 2.79E-08 | 0.149 |
| UC | 20 | 32104119 | rs6061159 | G | A | G | 0.064 | 0.010 | 1.77E-10 | 0.372 |
| UC | 20 | 32677356 | rs11699727 | A | A | G | 0.062 | 0.010 | 4.38E-10 | 0.371 |
| UC | 20 | 59249846 | rs259963 | A | A | C | 0.063 | 0.010 | 2.25E-10 | 0.440 |
| UC | 22 | 43162920 | rs6971 | A | A | G | -0.058 | 0.011 | 4.96E-08 | 0.342 |
| IBD | 1 | 6456955 | rs3007413 | T | T | C | 0.061 | 0.011 | 4.20E-08 | 0.158 |
| IBD | 1 | 186793684 | rs16825915 | T | T | C | -0.138 | 0.025 | 3.27E-08 | 0.029 |
| IBD | 2 | 202622300 | rs72926957 | G | C | G | 0.051 | 0.009 | 8.78E-09 | 0.305 |
| IBD | 2 | 236679892 | rs2720102 | A | A | G | -0.058 | 0.011 | 4.01E-08 | 0.173 |
| IBD | 3 | 42490447 | rs342518 | G | A | G | -0.049 | 0.009 | 1.22E-08 | 0.348 |
| IBD | 3 | 196688484 | rs75391517 | G | C | G | 0.061 | 0.010 | 1.36E-09 | 0.230 |
| IBD | 4 | 89015925 | rs2869974 | C | T | C | -0.060 | 0.011 | 1.28E-08 | 0.198 |
| IBD | 4 | 155679826 | rs4691843 | T | A | T | 0.052 | 0.008 | 4.78E-10 | 0.438 |
| IBD | 5 | 17118821 | rs1032763 | C | T | C | -0.049 | 0.009 | 1.56E-08 | 0.358 |
| IBD | 6 | 109054915 | rs11153158 | T | T | C | 0.076 | 0.013 | 1.79E-09 | 0.129 |
| IBD | 11 | 44593547 | rs4755847 | A | A | G | 0.047 | 0.008 | 1.72E-08 | 0.462 |
| IBD | 14 | 52401018 | rs2859837 | C | T | C | -0.050 | 0.009 | 3.25E-08 | 0.265 |
| IBD | 14 | 96172529 | rs56071894 | C | A | C | 0.068 | 0.011 | 8.83E-11 | 0.186 |
| IBD | 17 | 83073165 | rs11654537 | T | T | C | 0.052 | 0.009 | 2.40E-08 | 0.265 |
| IBD | 18 | 49646539 | rs72915967 | T | T | C | -0.062 | 0.009 | 1.95E-11 | 0.239 |
| IBD | 18 | 79456756 | rs77865749 | G | A | G | -0.061 | 0.011 | 3.34E-08 | 0.146 |
| IBD | 19 | 17068979 | rs2305754 | G | A | G | 0.054 | 0.009 | 1.84E-09 | 0.329 |
| IBD | 19 | 45811636 | rs7253302 | A | A | C | -0.047 | 0.009 | 4.06E-08 | 0.318 |
| IBD | 22 | 41389808 | rs28604950 | A | A | G | -0.057 | 0.010 | 2.56E-09 | 0.236 |

Variants are identified and annotated using the GRCh38 as the reference human genome sequence.

[a] Chr, chromosome.

[b] REF, reference allele in 1KGP EUR reference panel.

[c] Allele 1, Effect Allele.

[d] Allele 2, non-effect allele.

[e] SE, Standard error

[f] MAF, minor allele frequency mapped in 1KGP EUR reference panel.

* Variants that were previously reported in IBD.

**Table 2. Novel genetic risk variants identified in the European and East Asian GWAS meta-analysis.**

| Disease | Chr[a] | Position | Lead variant | REF[b] | Allele1[c] | Allele2[d] | Effect | SE[e] | *P* | MAF[f] |
|---|---|---|---|---|---|---|---|---|---|---|
| CD | 1 | 154319242 | rs1685633 | A | A | G | -0.101 | 0.018 | 4.04E-08 | 0.106 |
| CD | 2 | 173899186 | rs78098692 | C | T | C | -0.090 | 0.016 | 4.36E-08 | 0.169 |
| CD | 4 | 7049229 | rs28661210 | C | A | C | 0.063 | 0.012 | 3.89E-08 | 0.495 |
| CD | 5 | 156953056 | rs4704727 | T | T | G | 0.070 | 0.012 | 1.01E-08 | 0.331 |
| CD | 5 | 177372991 | rs10051765 | T | T | C | -0.069 | 0.012 | 3.01E-09 | 0.340 |
| CD | 6 | 3433106 | rs13206690 | T | T | G | 0.062 | 0.011 | 4.31E-08 | 0.378 |
| CD | 6 | 137698375 | rs72976823 | G | A | G | -0.067 | 0.012 | 4.17E-08 | 0.288 |
| CD | 10 | 6607973 | rs4748198 | C | T | C | -0.076 | 0.014 | 3.15E-08 | 0.200 |
| CD | 12 | 68110812 | rs1558744 | G | A | G | 0.065 | 0.012 | 2.51E-08 | 0.403 |
| CD | 17 | 56844977 | rs2176744 | G | A | G | 0.076 | 0.014 | 2.14E-08 | 0.291 |
| CD | 22 | 26642889 | rs5761655 | A | A | G | 0.066 | 0.012 | 1.93E-08 | 0.368 |
| UC | 1 | 209745158 | rs3817821 | A | A | T | 0.060 | 0.011 | 3.42E-08 | 0.233 |
| UC | 2 | 65452578 | rs1437461 | C | T | C | 0.061 | 0.010 | 6.68E-10 | 0.317 |
| UC | 2 | 119412310 | rs28478111 | C | T | C | 0.065 | 0.012 | 2.91E-08 | 0.230 |
| UC | 2 | 241832052 | rs79411665 | C | T | C | 0.092 | 0.015 | 7.91E-10 | 0.129 |
| UC | 4 | 101816990 | rs4411998 | T | T | C | -0.053 | 0.010 | 2.73E-08 | 0.377 |
| UC | 5 | 142866576 | rs59672579 | C | T | C | -0.103 | 0.016 | 1.78E-10 | 0.099 |
| UC | 8 | 129609008 | rs35389394 | C | T | C | -0.052 | 0.009 | 3.17E-08 | 0.426 |
| UC | 10 | 6608464 | rs1409874 | C | T | C | -0.067 | 0.011 | 5.43E-09 | 0.208 |
| UC | 10 | 88073835 | rs11202645 | G | A | G | 0.053 | 0.010 | 2.73E-08 | 0.423 |
| UC | 15 | 74935628 | rs4886640 | T | A | T | 0.064 | 0.012 | 3.12E-08 | 0.469 |
| UC | 16 | 75353680 | rs3851740 | T | T | C | 0.060 | 0.011 | 4.55E-08 | 0.439 |
| UC | 17 | 78681399 | rs7215260 | A | A | G | -0.054 | 0.010 | 1.16E-08 | 0.477 |
| IBD | 1 | 221262843 | rs9431466 | C | C | G | 0.047 | 0.008 | 2.11E-08 | 0.360 |
| IBD | 2 | 136302336 | rs10167166 | G | A | G | -0.065 | 0.012 | 3.80E-08 | 0.144 |
| IBD | 2 | 160487551 | rs7423572 | C | C | G | -0.053 | 0.010 | 3.00E-08 | 0.200 |
| IBD | 5 | 88830036 | rs304151 | T | T | G | 0.047 | 0.008 | 1.30E-08 | 0.268 |
| IBD | 5 | 134505105 | rs10079478 | T | T | C | -0.051 | 0.009 | 6.37E-09 | 0.248 |
| IBD | 8 | 58856561 | rs2726552 | G | A | G | -0.047 | 0.008 | 1.23E-09 | 0.464 |
| IBD | 16 | 80006686 | rs12921341 | G | T | G | -0.045 | 0.008 | 1.73E-08 | 0.469 |
| IBD | 17 | 78702868 | rs62075567 | C | A | C | 0.045 | 0.008 | 9.80E-09 | 0.486 |
| IBD | 19 | 49659652 | rs7251 | C | C | G | -0.046 | 0.008 | 2.79E-08 | 0.333 |

Variants are identified and annotated using the GRCh38 as the reference human genome sequence.

[a] Chr, chromosome.

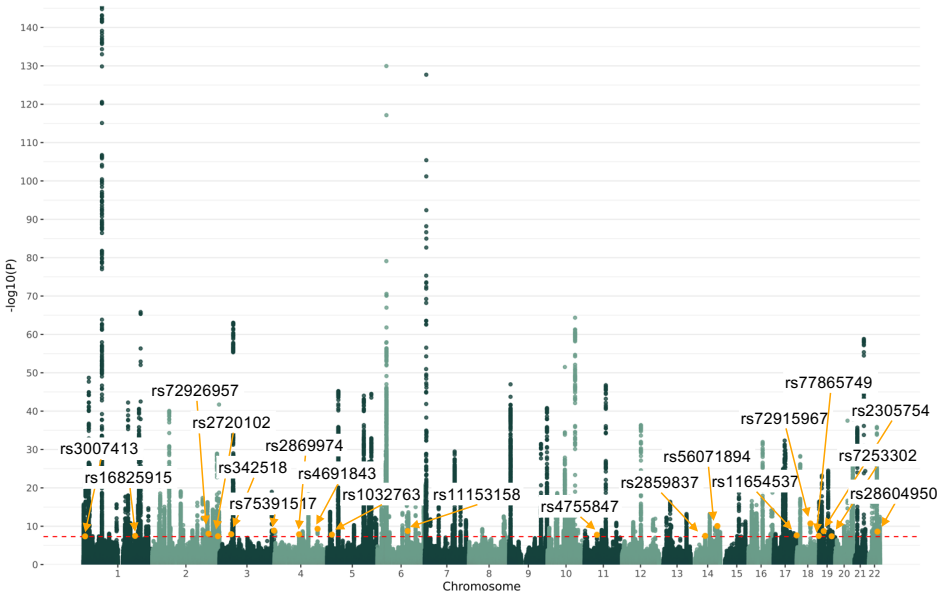
[b] REF, reference allele in 1KGP EUR reference panel.

[c] Allele 1, Effect Allele.

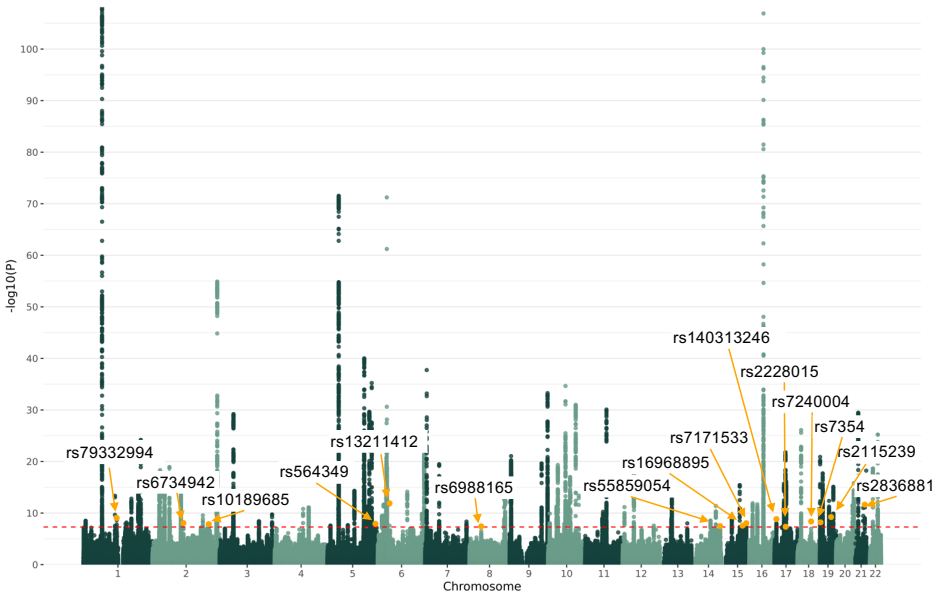
[d] Allele 2, non-effect allele.

[e] SE, Standard error

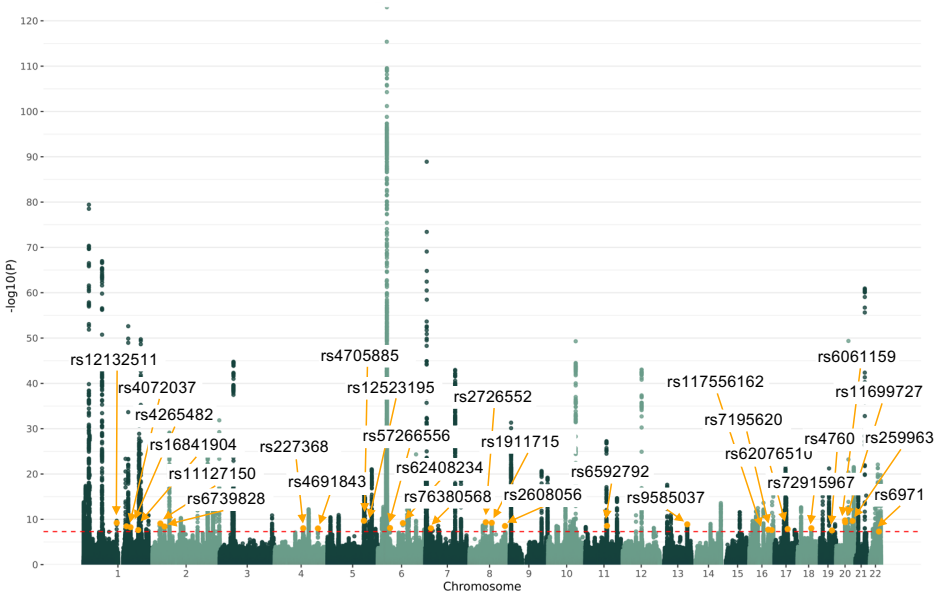
[f] MAF, minor allele frequency mapped in 1KGP EUR reference panel.



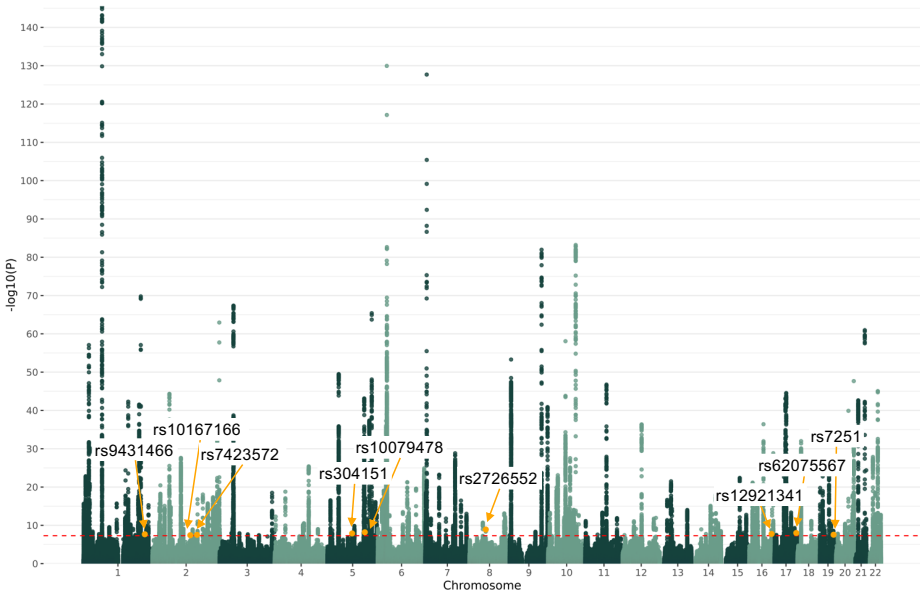

A. Inflammation Bowl Disease



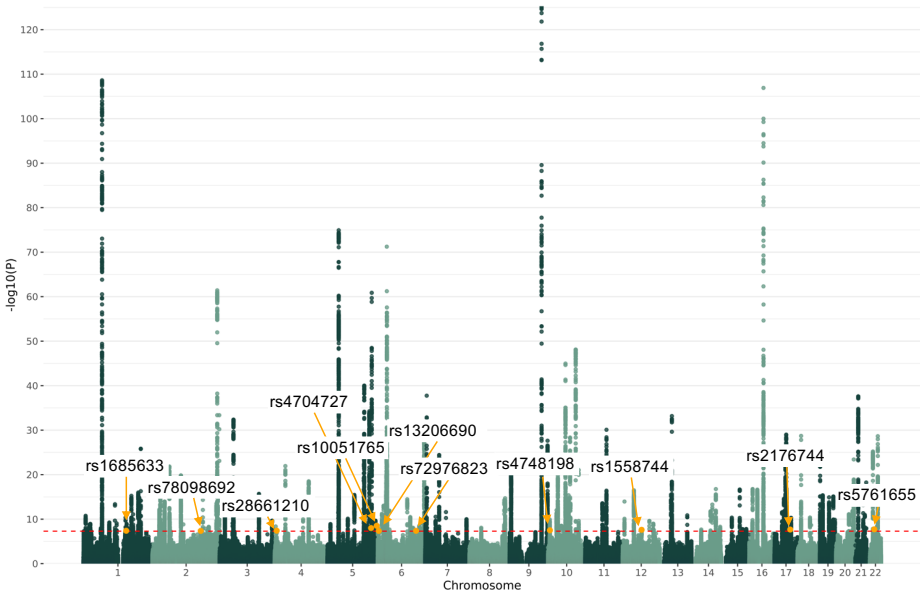

B. Crohn's disease



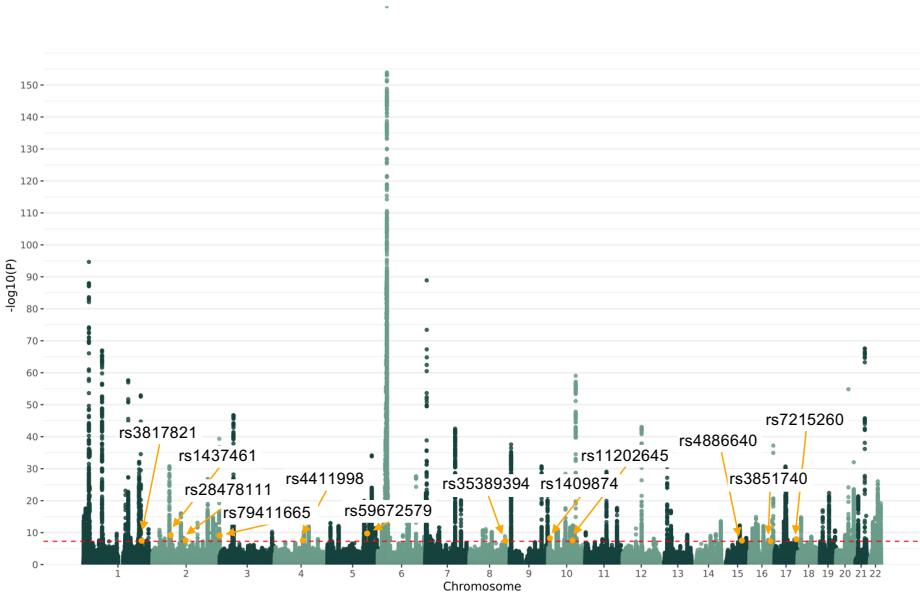

C. Ulcerative Colitis



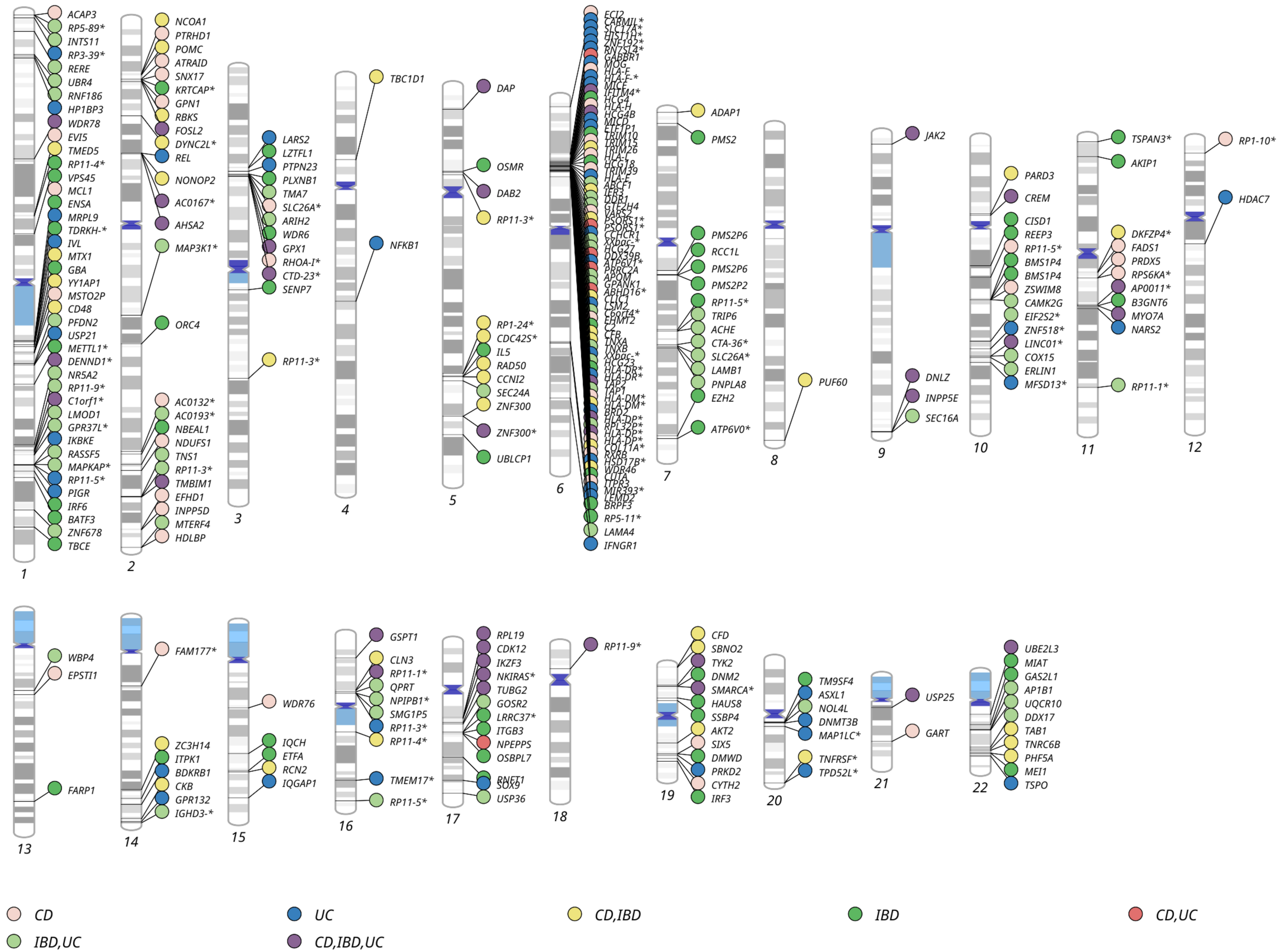

A. Inflammation Bowl Disease



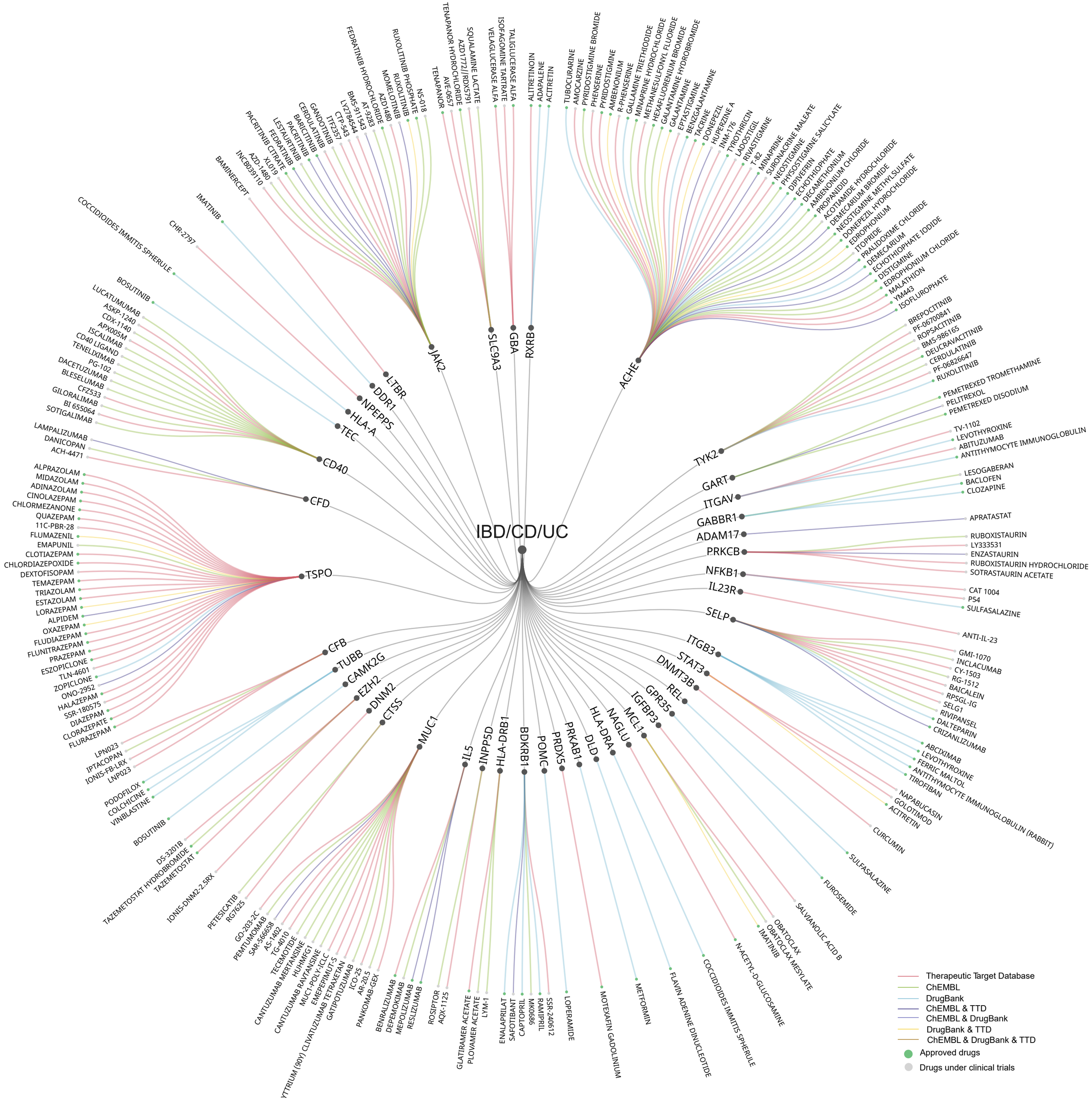

B. Crohn's disease




C. Ulcerative Colitis



1
ACAP3
RP5-89*
INTS11
RP3-39*
RERE
UBR4
RNF186
HP1BP3
WDR78
EVI5
TMED5
RP11-4*
VPS45
MCL1
ENSA
MRPL9
TDRKH-*
IVL
MTX1
GBA
YY1AP1
MSTO2P
CD48
PFDN2
USP21
METTL1*
DENND1*
NR5A2
RP11-9*
C1orf1*
LMOD1
GPR37L*
IKBKE
RASSF5
MAPKAP*
RP11-5*
PIGR
IRF6
BATF3
ZNF678
TBCE
2
NCOA1
PTRHD1
POMC
ATRAID
SNX17
KRTCAP*
GPN1
RBKS
FOSL2
DYNC2L*
REL
NONOP2
AC0167*
AHSA2
MAP3K1*
ORC4
AC0132*
AC0193*
NBEAL1
NDUFS1
TNS1
RP11-3*
TMBIM1
EFHD1
INPP5D
MTERF4
HDLBP
3
LARS2
LZTFL1
PTPN23
PLXNB1
TMA7
SLC26A*
ARIH2
WDR6
GPX1
RHOA-I*
CTD-23*
SENP7
RP11-3*
4
TBC1D1
NFKB1
5
DAP
OSMR
DAB2
RP11-3*
RP1-24*
CDC42S*
IL5
RAD50
CCNI2
SEC24A
ZNF300
ZNF300*
UBLCP1
6
ECI2
CARMIL*
SLC17A*
HIST1H*
ZNF192*
RN7SL4*
GABBR1
MOG
HLA-F
HLA-F-*
MICE
IFITM4*
HCG4
HLA-H
HCG4B
MICD
ETF1P1
TRIM10
TRIM15
TRIM26
HLA-L
HCG18
TRIM39
HLA-E
ABCF1
IER3
DDR1
GTF2H4
VARS2
PSORS1*
PSORS1*
CCHCR1
XXbac-*
HCG27
DDX39B
ATP6V1*
PRRC2A
APOM
GPANK1
ABHD16*
CLIC1
LSM2
C6orf4*
EHMT2
C2
CFB
TNXA
TNXB
XXbac-*
HCG23
HLA-DR*
HLA-DR*
TAP2
TAP1
HLA-DM*
HLA-DM*
BRD2
HLA-DP*
RPL32P*
HLA-DP*
HLA-DP*
COL11A*
RXRB
HSD17B*
WDR46
CUTA
ITPR3
MIR393*
LEMD2
BRPF3
RP5-11*
LAMA4
IFNGR1
7
ADAP1
PMS2
PMS2P6
RCC1L
PMS2P6
PMS2P2
RP11-5*
TRIP6
ACHE
CTA-36*
SLC26A*
LAMB1
PNPLA8
EZH2
ATP6V0*
8
PUF60
9
JAK2
DNLZ
INPP5E
SEC16A
10
PARD3
CREM
CISD1
REEP3
RP11-5*
BMS1P4
BMS1P4
ZSWIM8
CAMK2G
EIF2S2*
ZNF518*
LINC01*
COX15
ERLIN1
MFSD13*
11
TSPAN3*
AKIP1
DKFZP4*
FADS1
PRDX5
RPS6KA*
AP0011*
B3GNT6
MYO7A
NARS2
RP11-1*
12
RP1-10*
HDAC7
13
WBP4
EPSTI1
FARP1
14
FAM177*
ZC3H14
ITPK1
BDKRB1
CKB
GPR132
IGHD3-*
15
WDR76
IQCH
ETFA
RCN2
IQGAP1
16
GSPT1
CLN3
RP11-1*
QPRT
NPIPB1*
SMG1P5
RP11-3*
RP11-4*
TMEM17*
RP11-5*
17
RPL19
CDK12
IKZF3
NKIRAS*
TUBG2
GOSR2
LRRC37*
ITGB3
NPEPPS
OSBPL7
RNFT1
SOX9
USP36
18
RP11-9*
19
CFD
SBNO2
TYK2
DNM2
SMARCA*
HAUS8
SSBP4
AKT2
SIX5
DMWD
PRKD2
CYTH2
IRF3
20
TM9SF4
ASXL1
NOL4L
DNMT3B
MAP1LC*
TNFRSF*
TPD52L*
21
USP25
GART
22
UBE2L3
MIAT
GAS2L1
AP1B1
UQCR10
DDX17
TAB1
TNRC6B
PHF5A
MEI1
TSPO
CD
IBD,UC
UC
CD,IBD,UC
CD,IBD
IBD
CD,UC


IBD/CD/UC
JAK2
SLC9A3
GBA
RXRB
ACHE
LTBR
DDR1
NPEPPS
HLA-A
TEC
CD40
CFD
TSPO
TYK2
GART
ITGAV
GABBR1
ADAM17
PRKCB
NFKB1
IL23R
SELP
ITGB3
STAT3
DNMT3B
REL
GPR35
IGFBP3
MCL1
NAGLU
HLA-DRA
DLD
PRKAB1
PRDX5
POMC
BDKRB1
HLA-DRB1
INPP5D
IL5
MUC1
CTSS
DNM2
EZH2
CAMK2G
TUBB
CFB
TENAPANOR
TENAPANOR HYDROCHLORIDE
AVE-0657
AZD1772//RDX5791
SQUALAMINE LACTATE
VELAGLUCERASE ALFA
ISOFAGOMINE TARTRATE
TALIGLUCERASE ALFA
ALITRETINOIN
ADAPALENE
ACITRETIN
NS-018
RUXOLITINIB PHOSPHATE
RUXOLITINIB
MOMELOTINIB
AZD1480
FEDRATINIB HYDROCHLORIDE
AT-9283
BMS-911543
LY2784544
CTP-543
ITF2357
GANDOTINIB
CERDULATINIB
BARICITINIB
PACRITINIB
FEDRATINIB
LESTAURTINIB
PACRITINIB CITRATE
XL019
AZD-1480
INCB039110
BAMINERCEPT
IMATINIB
CHR-2797
COCCIDIOIDES IMMITIS SPHERULE
BOSUTINIB
LUCATUMUMAB
ASKP-1240
CDX-1140
APX005M
ISCALIMAB
CD40 LIGAND
TENELIXIMAB
PG-102
DACETUZUMAB
BLESELUMAB
CFZ533
GILORALIMAB
BI 655064
SOTIGALIMAB
LAMPALIZUMAB
DANICOPAN
ACH-4471
ALPRAZOLAM
MIDAZOLAM
ADINAZOLAM
CINOLAZEPAM
CHLORMEZANONE
QUAZEPAM
11C-PBR-28
FLUMAZENIL
EMAPUNIL
CLOTIAZEPAM
CHLORDIAZEPOXIDE
DEXTOFISOPAM
TEMAZEPAM
TRIAZOLAM
ESTAZOLAM
LORAZEPAM
ALPIDEM
OXAZEPAM
FLUDIAZEPAM
FLUNITRAZEPAM
PRAZEPAM
ESZOPICLONE
TLN-4601
ZOPICLONE
ONO-2952
HALAZEPAM
SSR-180575
DIAZEPAM
CLORAZEPATE
FLURAZEPAM
LPN023
IPTACOPAN
IONIS-FB-LRX
LNP023
PODOFILOX
COLCHICINE
VINBLASTINE
BOSUTINIB
DS-3201B
TAZEMETOSTAT HYDROBROMIDE
TAZEMETOSTAT
IONIS-DNM2-2.5RX
PETESICATIB
RG7625
GO-203-2C
PEMTUMOMAB
SAR-566658
AS-1402
TG-4010
TECEMOTIDE
CANTUZUMAB MERTANSINE
HUHMFG1
CANTUZUMAB RAVTANSINE
MUC1-POLY-ICLC
EMEPEPIMUT-S
YTTRIUM (90Y) CLIVATUZUMAB TETRAXETAN
GATIPOTUZUMAB
ICO-25
AR-20.5
PANKOMAB-GEX
BENRALIZUMAB
DEPEMOKIMAB
MEPOLIZUMAB
RESLIZUMAB
ROSIPTOR
AQX-1125
GLATIRAMER ACETATE
PLOVAMER ACETATE
LYM-1
ENALAPRILAT
SAFOTIBANT
CAPTOPRIL
WIN0396
RAMIPRIL
SSR-240612
LOPERAMIDE
MOTEXAFIN GADOLINIUM
METFORMIN
FLAVIN ADENINE DINUCLEOTIDE
COCCIDIOIDES IMMITIS SPHERULE
N-ACETYL-D-GLUCOSAMINE
IMATINIB
OBATOCLAX
OBATOCLAX MESYLATE
SALVIANOLIC ACID B
FUROSEMIDE
SULFASALAZINE
CURCUMIN
ACITRETIN
GOLOTIMOD
NAPABUCASIN
TIROFIBAN
ANTITHYMOCYTE IMMUNOGLOBULIN (RABBIT)
FERRIC MALTOL
LEVOTHYROXINE
ABCIXIMAB
CRIZANLIZUMAB
DALTEPARIN
RIVIPANSEL
SELG1
RPSGL-1G
BAICALEIN
RG-1512
CY-1503
INCLACUMAB
GMI-1070
ANTI-IL-23
SULFASALAZINE
CAT 1004
P54
RUBOXISTAURIN
LY333531
ENZASTAURIN
RUBOXISTAURIN HYDROCHLORIDE
SOTRASTAURIN ACETATE
APRATASTAT
LESOGABERAN
BACLOFEN
CLOZAPINE
TV-1102
LEVOTHYROXINE
ABITUZUMAB
ANTITHYMOCYTE IMMUNOGLOBULIN
PEMETREXED TROMETHAMINE
PELITREXOL
PEMETREXED DISODIUM
BREPOCITINIB
PF-06700841
ROPSACITINIB
BMS-986165
DEUCRAVACITINIB
CERDULATINIB
PF-06826647
RUXOLITINIB
TUBOCURARINE
AMOCARZINE
PYRIDOSTIGMINE BROMIDE
PHENSERINE
PYRIDOSTIGMINE
AMBENONIUM
R-PHENSERINE
GALLAMINE TRIETHIODIDE
MINAPRINE HYDROCHLORIDE
METHANESULFONYL FLUORIDE
HEXAFLUORENIUM BROMIDE
GALANTAMINE HYDROBROMIDE
GALANTAMINE
EPTASTIGMINE
BENZGALANTAMINE
TACRINE
DONEPEZIL
HUPERZINE A
INM-176
TYROTHRICIN
LADOSTIGIL
RIVASTIGMINE
T-82
MINAPRINE
SURONACRINE MALEATE
NEOSTIGMINE
PHYSOSTIGMINE SALICYLATE
DIPIVEFRIN
ECHOTHIOPHATE
DECAMETHONIUM
AMBENONIUM CHLORIDE
PROPANIDID
ACOTIAMIDE HYDROCHLORIDE
DEMECARIUM BROMIDE
NEOSTIGMINE METHYLSULFATE
DONEPEZIL HYDROCHLORIDE
EDROPHONIUM
ITOPRIDE
PRALIDOXIME CHLORIDE
DEMECARIUM
ECHOTHIOPHATE IODIDE
DISTIGMINE
EDROPHONIUM CHLORIDE
MALATHION
YM443
ISOFLUROPHATE
Therapeutic Target Database
ChEMBL
DrugBank
ChEMBL & TTD
ChEMBL & DrugBank
DrugBank & TTD
ChEMBL & DrugBank & TTD
Approved drugs
Drugs under clinical trials