Multiomics-based causal inference identifies novel therapeutic targets for inflammatory bowel disease in East Asians

Jung Oh Kim (✉ kjo@basgenbio.com)
Basgenbio, Inc.

Sun Yeop Lee
Basgenbio, Inc.   https://orcid.org/0000-0001-9332-049X

Eun Gyo Kim
Basgenbio, Inc.

Eunbyeol Lee
Basgenbio, Inc.

Jihye Lee
Basgenbio, Inc.   https://orcid.org/0009-0008-6608-5859

Yu Jeong Shim
Basgenbio, Inc.   https://orcid.org/0000-0003-0138-9567

Article

Additional Declarations: Yes there is potential Competing Interest. S.Y.L., E.K., E.L., J.L., Y.J.S., and J.O.K. are employees of Basgenbio, Co., Seoul, Republic of Korea. All authors declare no other competing interests.

Abstract

Multiomics-based efforts to identify therapeutic targets for IBD have been limited to European populations. Prior reports on heterogeneity between East Asians and Europeans in clinical manifestations of IBD and genetic architectures of IBD-related variants warrant a separate investigation in East Asians. Using the East Asian genome and proteome data, we applied two multiomics-based causal inference methods, proteome-wide Mendelian randomization and causal proteome-wide association study. For IBD, Crohn's disease (CD), and ulcerative colitis (UC), we found 30 potential drug targets with proteomic evidence. IL18R1, IL1RL1, KIR3DL1, and MEP1B had consistent associations with across IBD, CD, and UC. Fifteen targets were CD-specific, while eight were UC-specific. Among the candidate targets, thirteen and eight had supportive MR evidence in the plasma transcriptome data and the multi-tissue transcriptome data of European ancestry, respectively. IL18R1, IL6R, IL16, TNFRSF14 or their direct interactors were currently targeted by drugs being developed to treat IBD. IL1RL1 and PDGFRB had existing drugs that may be repurposed for IBD. Crucially, we identified six previously unreported target genes, opening new avenues for therapeutic interventions in IBD that warrant immediate validation in upcoming experiments and clinical trials.

1. Introduction

Inflammatory bowel disease (IBD), of which two main types are Crohn's disease (CD) and ulcerative colitis (UC), is a category of conditions with chronic inflammation of the gastrointestinal tract. IBD causes severe abdominal pain and diarrhea, and potentially leads to a range of sequelae including colon cancer. Globally, over 6 million individuals suffer from IBD, which now ranks as the fourth-leading cause of years lived with disability among digestive diseases[1]. While incidence rates of IBD in North America, Europe, Australia have fallen or stabilized since the 1990s, countries in Asia, Africa, and South America have experienced an increase in IBD incidence and prevalence, imposing substantial burden on the economy and especially on the healthcare systems.

One notable observation of IBD in addition to its diverging epidemiological trends is the potential heterogeneity in its pathogenesis across ethnic or ancestral populations, especially between East Asians and Europeans. Population-based cohort studies found that East Asian patients with IBD were more likely to exhibit perianal involvement, complicated progression, extraintestinal manifestations, and male predominance, after adjusting for a range of socioeconomic and demographic factors[2,3]. Genome-wide association studies (GWAS) reported substantial differences of important IBD-related variants between East Asians and Europeans in their minor allele frequencies and effect sizes, suggesting potentially distinct genetic influences on IBD development in East Asians[4–7].

Currently, there is no cure for IBD. Its etiology is understood to be multifaceted with genetic and environmental factors, although the exact cause of IBD has not been pinpointed[8]. Current treatment regimens aim to control inflammation and other symptoms, but they are far from ideal as clinical

remission rates remain low and unstable over time[9]. Complexities in pathophysiological mechanisms of IBD present significant challenges in developing effective drugs.

Recently, increasing availability of multiomics data coupled with explicit causal inference techniques such as Mendelian randomization (MR) enabled multidimensional approaches that have proven to be effective for discovering novel therapeutic targets and pathways[10-13]. For IBD, various data sources including genome, transcriptome, epigenome, proteome, metabolome, and gut microbiome have been interrogated[14-20]. However, most evidence on the IBD multiomics is currently limited to data of European ancestry. For example, a previous proteome-wide screening of IBD drug targets has identified MST1, HGFAC, STAT3, ITPKA, and CXCL5 as potential novel drug targets for IBD or UC, but only utilized European data[16]. A major constraint has been that, unlike genomic data which has increasingly become diverse, multiomics data are still mostly limited to Europeans.

Unique proteome data of East Asian individuals were released a few months ago, and a latest large-scale IBD GWAS was conducted in East Asians, enabling an exploration of IBD multiomics specifically in this population. In this study, therefore, we applied two approaches of multiomics-based causal inference to screen for potential therapeutic target genes for IBD and its subtypes (CD, UC) using East Asian data (Fig. 1). First, we performed proteome-wide Mendelian randomization (MR) which indirectly circumvents confounding under an instrumental variable (IV) framework. Second, we calculated cis-genetic effects on proteins and used them to directly adjust for confounding via a summary-based proteome-wide association study (PWAS) with its novel extension, causal-PWAS. Third, to provide additional evidence and aid in interpretation for identified candidate targets, we carried out proteome-wide colocalization, transcriptome-wide MR, phenome-wide association study (PheWAS) database scan, functional enrichment analyses, and drug-target database search.

2. Results

2.1. Proteome-wide MR

Among 1464 proteins, 134 (9.15%) were successfully analyzed for proteome-wide MR. Of the proteins that were not analyzed, two were excluded because pQTLs were absent in the IBD GWAS, one was excluded because pQTLs were absent in 1KGP for clumping, and 1,324 were excluded because they did not have statistically significant pQTLs in the East Asian samples. For IBD, proteins with MR-cML estimates with FDR-corrected $p < 0.05$ were CNTN2, CSTB, FGF5, ICAM5, IL18R1, IL1RL1, KIR3DL1, MEP1B, PDCD6, PDGFRB, PILRA, PILRB, PM20D1, and PNLIPRP2 (Fig. 2, Supplementary Table 1). For CD, CD300C, CNTN2, IL16, IL18R1, IL1RL1, IL6R, KIR3DL1, LILRA2, MEP1B, OBP2B, PILRA, PILRB, SPINK5, and TNXB, whereas for UC, CSTB, IL18R1, IL1RL1, KIR3DL1, MEP1B, PDCD6, PDGFRB, and PNLIPRP2 had statistically significant associations. The correlation between CD and UC MR estimates was statistically significant (r = 0.588, p < 0.001; Supplementary Fig. 1).

2.2. PWAS & Causal-PWAS extension

Among 487,199 UKB participants with genotype data, 54,219 individuals had protein data. Our random forest classifier assigned East Asian ancestry for 230 individuals, a number that closely matched the number reported in a previous publication[21] (Supplementary Table 2). In the PWAS, 258 proteins (17.62%) among 1464 proteins were successfully analyzed, while the rest did not achieve statistically significant SNP-based heritability ($p < 0.01$). For IBD, EFNA1, IL18R1, KLK12, KLK13, RNASET2, RSPO1, SUSD2 had causal-PWAS PIP > 0.80 (Fig. 3, Supplementary Table 3). For CD, CD300C, EFNA1, IL18R1, RNASET2, SIRPB1, SUSD2, and TNFRSF14, whereas for UC, HYAL1, IL18R1, KLK12, KLK13, and RSPO1 had causal-PWAS PIP > 0.80. The correlation between CD and UC PWAS estimates was statistically significant ($r = 0.356$, $p < 0.001$; Supplementary Fig. 1).

2.3. Multiomics evidence integration

When combining all genomic and proteomic evidence from the two causal inference methods (proteome-wide MR and causal-PWAS), we identified a total of 30 proteins causally associated with the risk of IBD or its subtypes (Table 1). According to the drug-target and PheWAS databases, 24 of these proteins (80%) are known targets. A higher protein level of IL18R1, IL1RL1, or MEP1B, or a lower level of KIR3DL1 was associated with a higher risk for all three disease types (i.e., IBD-encompassing genes). Among these, KIR3DL1 and MEP1B also had coloc PP-H_3 < 0.80 (Supplementary Table 4). The association for KIR3DL1 was also observed for IBD and UC in transcriptome-wide MR analyses in both the plasma and multi-tissue data (Supplementary Table 5, 6). IL1RL1 had the transcriptomic association with CD in the plasma tissue data and with UC in the multi-tissue data. A higher level of FGF5, ICAM5, or PM20D1 was associated with a higher risk of IBD and had coloc PP-H_3 < 0.80, but the associations were not observed when separately analyzed in the CD or UC data.

Table 1 Evidence integration for druggable target identification.

Target group	Gene	pQTL–MR	Causal-PWAS	Novelty
IBD-encompassing (= IBD and CD and UC)	IL18R1	↑	↑	.
	IL1RL1	↑	—	.
	KIR3DL1	↓	—	.
	MEP1B	↑	—	.
CD-specific (= CD only or with IBD)	CD300C	↑	↑	.
	CNTN2	↑	—	✓
	EFNA1	.	↓	.
	IL16	↓	.	.
	IL6R	↑	.	.
	LILRA2	↑	—	.
	OBP2B	↓	—	✓
	PILRA	↑	—	.
	PILRB	↑	—	.
	RNASET2	.	↓	.
	SIRPB1	.	↓	.
	SPINK5	↑	—	.
	SUSD2	.	↓	.
	TNFRSF14	.	↓	.
	TNXB	↑	—	.
UC-specific (= UC only or with IBD)	CSTB	↑	—	✓
	HYAL1	.	↓	.
	KLK12	—	↓	.
	KLK13	.	↑	✓
	PDCD6	↓	.	✓
	PDGFRB	↓	.	.
	PNLIPRP2	↓	—	✓
	RSPO1	.	↓	.
IBD-specific (= IBD only but not separately with CD and UC)	FGF5	↑	.	.
	ICAM5	↑	.	.
	PM20D1	↑	.	.

For CD specifically, and not UC, a higher protein level of CD300C, CNTN2, IL6R, LILRA2, PILRA, PILRB, SPINK5, or TNXB, or a lower level of EFNA1, IL16, OBP2B, RNASET2, SIRPB1, SUSD2, or TNFRSF14 was associated with a higher risk. All but IL6R, LILRA2, and TNXB had coloc PP-H_3 < 0.80, and IL16 had coloc PP-H_4 > 0.80 in addition. Among the CD-specific genes, CNTN2, EFNA1, PILRA, PILRB, RNASET2, and SUSD2 also had associations with IBD as a whole. At the transcriptome level, CD300C and LILRA2 had associations with CD in both the plasma and multi-tissue data. IL6R, PILRA, and RNASET2 had

associations with CD in the plasma tissue data only, and IL16 had associations with CD in the multi-tissue data only.

A higher level of CSTB or KLK13, or a lower level of HYAL1, KLK12, PDCD6, PDGFRB, PNLIPRP2, or RSPO1 was associated with a higher risk of UC, but not CD. All of these UC-specific genes passed the coloc PP-H$_3$ < 0.80 threshold. Except HYAL1, all also had associations with IBD as a whole. At the transcriptome level, PDCD6 also had associations with UC in the plasma tissue data, while CSTB had associations with UC in both the plasma and multi-tissue data.

2.4. Gene-level PheWAS database scan

Among 30 candidate targets, the European pQTL MR database also included the associations between IL18R1, IL1RL1, and ICAM5 with IBD and CD, but not UC (Supplementary Table 7). In addition, associations were found for IL18R1 with eczema, CNTN2 for mean platelet volume, and IL6R for rheumatoid arthritis. The rare variant aggregation results showed that all candidate target genes did not have statistically significant pleiotropic associations in other phenotypes, except for MEP1B, PILRA, and PM20D1, which had associations with protein levels of MEP1A, PILRB, and TOP1, respectively. These proteins did not have further pleiotropic associations with other phenotypes.

2.5. Functional enrichment analysis

GO enrichment analyses showed that among the biological process terms, the candidate targets were enriched in positive regulation of cytokine production (IL16, IL18R1, IL1RL1, IL6R, LILRA2, TNFRSF14, TNXB) (Supplementary Fig. 2). Among the molecular function terms, they were also enriched in immunity-related pathways such as cytokine binding, immune receptor activity, growth factor receptor binding (IL18R1, IL1RL1, IL6R, TNFRSF14, KIR3DL1, LILRA2, PILRA, PILRB, HYAL1, FGF5, PDGFRB). Among the cellular component terms, only lysosomal lumen was significantly enriched (HYAL1, PDGFRB, RNASET2). Enriched KEGG pathways and modules also included immunity-related pathways such as cytokine-cytokine receptor interaction.

2.6. Druggable target identification

Among the identified potential drug targets, two were direct targets of drugs with approval, four were in a clinical trial stage, five were in an experimental stage, sixteen were not currently in development but annotated as druggable, and four were not annotated as druggable (Supplementary Table 8). IL6R had three drugs with regulatory approvals and two in development, and their indications included immune-related diseases such as rheumatoid arthritis, coronavirus disease 19 (COVID-19), and systemic lupus erythematosus. Notably, one drug, olamkicept, was successful in a phase 2 clinical trial for UC. Its ligand, IL6, was also a target of three drugs that have been in development for treating CD (Supplementary Table 9). PDGFRB was targeted by a number of approved drugs or drugs in development with indications ranging from idiopathic pulmonary fibrosis to various types of cancer. IL18R1 is targeted by iboctadekin which completed a phase 2 clinical trial for treating melanoma. Also, IL18, which binds to IL18R1, is targeted by GSK-1070806 which completed a phase 2 clinical trial for CD. IL1RL1 is targeted by

astegolimab which has recently started a phase 3 clinical trial for chronic obstructive pulmonary disease and completed phase 2 clinical trials for COVID-19, asthma, and atopic eczema. While IL16 and TNFRSF14 are still in an experimental stage, IL16 receptors (GRIN2A, CCR5) and a ligand of TNFRSF14 (TNFSF14) have drugs for CD or UC in a clinical trial stage.

3. Discussion

Using the East Asian genome and proteome data, we identified potential therapeutic targets for IBD and its subtypes. We applied two multiomics-based causal inference methods, proteome-wide MR and causal-PWAS and found 30 potential drug targets with proteomic evidence Four of them (IL18R1, IL1RL1, KIR3DL1, and MEP1B) had consistent associations with the risk of IBD, CD, and UC. Fifteen targets were CD-specific, while eight were UC-specific. Among the candidate targets, thirteen and eight had supportive MR evidence in the plasma transcriptome data and the multi-tissue transcriptome data of European ancestry, respectively. IL18R1, IL6R, IL16, and TNFRSF14 or their direct interactors were currently targeted by drugs being developed to treat IBD. IL1RL1 and PDGFRB had existing drugs that may be repurposed for IBD. Six target genes were novel according to a wide range of drug-target and PheWAS databases.

IBD multiomics had been largely limited to European ancestry. However, a prior IBD GWAS in East Asians revealed substantial heterogeneity in minor allele frequencies and effect sizes for several IBD-related variants, implying significant downstream consequences for multiomics-based therapeutic target discovery[4]. Indeed, the potential IBD drug targets we identified in East Asians were mostly not detected by a similar previous study in Europeans[16], even when some of these are known IBD targets and have been developed for drugs. Further research on these novel targets may reveal mechanisms that could explain previously reported ethnic differences in clinical manifestations of IBD[2,3,22]. Using data of diverse ancestries in multiomics is fruitful, and data collection in transcriptome, proteome, microbiome, and other omics data should follow a current effort to increase diversity in genomic data. Furthermore, a recent review on IBD drug randomized controlled trials (RCT) reported a low representation of non-White participants (14.7%)[23]. As a societal burden of IBD is expected to rise in geographical regions such as Asia, Africa, and South America[1], an effort to develop optimal IBD therapeutics should diversify its participants in every step of the entire process from multiomics-based screening to confirmatory RCTs.

Several of our candidate targets have established relationships with IBD and are direct targets of existing drugs. Some are already being developed to treat IBD, while others may be promising for drug repurposing. IL18R1 and IL1RL1 are cytokine receptors within the IL1 receptor family which, together with IL1 family, is an established regulator of gastrointestinal inflammation. IL18R1 binds to IL18 and is essential for IL18-mediated signal transduction. Elevated levels of IL18 in gut epithelium have been linked to mucosal barrier breakdown and intestinal integrity[24–26]. IL18R1 has been targeted by iboctadekin which was developed to treat melanoma and non-Hodgkin's lymphoma. GSK-1070806, an anti-IL18 monoclonal antibody, is being developed to treat a range of conditions including IBD, atopic eczema, and type 2 diabetes[27]. IL1RL1 is a receptor for IL33, and studies have suggested relationships between

IL33/IL1RL1 signaling and colitis, tissue fibrosis, and mucosa healing[28–30]. Astegolimab inhibits IL1RL1 to treat asthma, atopic eczema, chronic obstructive pulmonary disease, and COVID-19. Previous MR studies using European data reported associations between IBD and the IL1 or IL1 receptor family proteins, and our studies now provided consistent evidence in East Asians[15,17,31,32].

IL6R is a subunit of IL6 receptor complex, and through IL6/IL6R axis, sustained elevation of IL6 conveys IL6 classic and trans-signaling, which is primarily responsible for chronic inflammation. Elevated levels of IL6 have been observed in IBD[33,34], and animal studies showed that IL6 inhibition was effective in CD[35,36]. Olamkicept, a selective inhibitor of the IL6/sIL6R complex, was developed to inhibit IL6 trans-signaling without blocking IL6 classic signaling, and a phase 2 clinical trial showed its efficacy in clinical response for UC patients[37]. Several other IL6 inhibitors (olokizumab, clazakizumab, PF-04236921) are also being tested against CD. Existing drugs targeting IL6R (e.g., tocilizumab, sarilumab, satralizumab) have been approved for multiple immune-related indications including rheumatoid arthritis; these and other IL6R inhibitors in trials (e.g., levilimab, vobarilizumab) presents promising opportunities for repurposing against IBD.

One interesting case is TNFRSF14 of which its ligand, TNFSF14, was tested as an IBD drug. Quisovalimab, a TNFSF14 inhibitor, had two phase 1 IBD clinical trials prematurely terminated and one phase 2 trial that showed efficacy for non-eosinophilic asthma in patients with high baseline TNSFSF14 levels. TNFSF14 inhibition may be appropriate considering its traditional role in proinflammation, but our MR results suggested increased levels of TNFRSF14 were associated with a lower risk of IBD in both proteomic and transcriptomic data. This is consistent with a study that found TNFSF14-deficient mice having more serious colitis than wild-type mice, which illuminated a separate unexpected role of TNFSF14 in protecting against intestinal inflammation[38].

We found multiomics evidence for several novel therapeutic targets and some weakly linked candidate targets for IBD. RSPO1 is an activator of the canonical Wnt signaling pathway and is highly expressed in various tissues in the gastrointestinal tract. It has a crucial role in stem cell homeostasis in the gastrointestinal tract[39], and in vitro studies showed that RSPO1 supported the survival of intestinal stem cells, while its deficiency resulted in crypt cell and apoptosis[40]. KLK12 and KLK13 are kallikreins which have functions in carcinogenesis and are close interactors of SPINK5, another candidate target from our analyses. A structural network of IL10 centered signaling found that interactions of KLK13 may lead to cleavage in the major components of extracellular matrix and promote cancer cell growth and metastasis[41]. SPINK5 is a serine protease inhibitor with functions in anti-inflammatory and anti-microbial protection of mucous epithelia, and contributes to integrity and protective barrier function of the skin. SPINK5, KLK12, and KLK13 gene and protein expression are all highly expressed in mucosa and esophagus tissues, and for SPINK5, also in colon and rectum tissues. Mutations in SPINK5 promotes non-specific inflammation independent of allergens, leading to pro-T helper responses related to signaling molecules such as IL33, a target of immune-related diseases such as asthma and atopic eczema[42]. SPINK5 has also been suggested as a therapeutic target for cancer[43]. Other novel candidates, CSTB,

PDCD6, and PDGFRB, had limited research regarding IBD. They are all highly expressed in mucosal tissues for mRNA or protein. Weak links have been found between CSTB and UC neoplastic progression[44], and between PDCD6 and infliximab responses and fibrosis in CD[45,46]. Upregulation of PDCD6 interacting protein (PDCD6IP) has also been linked with UC[47]. PDGFRB has several approved drugs with indications including colorectal cancer and gastrointestinal stromal tumor which are major sequelae of IBD, and it forms heterodimers with PDGFRA which has been linked with colitis susceptibility[48] and postnatal intestinal epithelial stemness[49]. In fact, nintedanib, an approved PDGFRB inhibitor for idiopathic pulmonary fibrosis and systemic scleroderma, has been found to inhibit an mRNA level of EFNA1, another candidate target we found, to alleviate colitis and improved a gut microbiota composition in a mouse model[50].

There were important limitations to note. First, the UKB proteome data provided only a limited coverage of the entire plasma proteome. For example, among 66 targets of approved IBD drugs according to Open Target Platform, only six were included in the data, and only one had variants with sufficient strength to perform MR. Second, the proteome data was only quantified by antibody-based assays. Distinct and complementary nature of antibody- and aptamer-based assays have been reported[51], and it would be consequential for drug target discovery, but aptamer-based proteome data was not available for East Asians. Third, the sample size of East Asians in the proteome data was small. Low statistical power for pQTLs particularly tamed our interpretation of colocalization, which requires high statistical power. Sample size was small (less than 1000) across all ancestries except for the European ancestry; increasing diversity in proteomic data is warranted. Fourth, our transcriptomic analyses relied on European data because large, publicly available transcriptomic data of East Asians do not exist. We interpreted the transcriptomic findings with caution.

Employing a set of causal inference methods on East Asian data, we provided proteomic evidence for several potential therapeutic targets for IBD and supportive findings from transcriptomic analyses, PheWAS, and functional enrichment analyses. Some targets or their direct interactors (IL18R1, IL6R, IL16, TNFRSF14) had existing drugs that were currently being developed to treat IBD; we provided additional evidence supporting these developments. Other targets (IL1RL1, PDGFRB) had drugs for other indications; these may be promising candidates for IBD drug repurposing. Several novel targets were found, which may be attributed to the use of East Asian data and the application of causal-PWAS. They should be validated in future experiments and trials.

4. Methods
4.1. IBD GWAS summary data

East Asian genome-wide association study (GWAS) summary statistics for IBD and its subtypes were obtained from International Inflammatory Bowel Disease Genetics Consortium (IIBDGC). IIBDGC has recently conducted the largest IBD GWAS meta-analysis of East Asian ancestry with 14,393 cases (CD:

7,372; UC: 6,682) and 15,456 controls[4]. The meta-analysis included cohorts of individuals from China, Korea, Japan, and Hong Kong. Further details of the included cohorts, genotype quality control procedures, and GWAS procedures are described in the original publication[4].

4.2. Proteome data

Proteome data are appropriate as a primary data source for target discovery because proteins are more genetically proximal, have relatively distinct genetic architectures compared to more distal biomarkers with polygenic architectures, and are direct targets of most drugs, including small molecules and biologics.

East Asian protein data were obtained from UK Biobank (UKB) in two forms: individual-level protein data and pQTL summary statistics[52]. UKB is a population-based cohort of approximately 500,000 individuals recruited between 2006 and 2010 in the UK. Participants' information has been collected via a wide range of sources such as electronic health records, death and cancer registries, self-reported surveys, biomarker measurements, and medical images. In 2023, proteomic profiling data were released as part of the UKB Pharma Proteomics Project (UKB-PPP). Brief information on proteome data is described below, and further details on proteomic profiling and GWAS procedures are described in an original publication[52].

Individual-level protein data were obtained from the UK Biobank with an approved application number 77890. The UK Biobank obtained approval from the Northwest Multicenter Research Ethics Committee. All participants provided written informed consent. Using the Olink Explore 1536 platform, proteomic profiling was conducted on blood plasma samples of 54,219 individuals (46,595 randomly selected samples, 6,376 selected by UKB-PPP consortium member companies, and 1,268 selected from the UKB COVID-19 repeat imaging study; 20 individuals overlapping in the consortium-selected and the imaging study). Baseline demographic and clinical characteristics are described elsewhere[21]. Data on 1,463 unique proteins were available at the time of this study. Each protein level was quantified in a normalized protein expression unit, where 1 unit difference represents a doubling of protein concentration.

The UKB East Asian pQTL summary statistics were publicly available. The pQTL GWAS were generated for 262 East Asian individuals using REGENIE v2.2.1, which adjusts for population structure[53]. Variants were included based on the following criteria: minor allele frequency (MAF) > 0.01, minor allele count > 100, genotyping rate > 0.99, Hardy-Weinberg equilibrium (HWE) $p > 1 * 10^{-15}$, missingness < 0.10, linkage disequilibrium (LD) pruning ($r^2 < 0.8$)[52].

4.3. Transcriptome data

A large, plasma tissue transcriptome data source, the eQTLGen Consortium, and a multi-tissue transcriptome data source, the GTEx Consortium, were used for transcriptomic analyses. The eQTLGen, cis-eQTL (± 1 MB) summary statistics were produced with 31,684 blood samples from meta-analyses of 37 consortium cohorts[54]. GWASs were conducted in each cohort, and their Z scores were weighted by

sample size. SNPs with MAF > 0.01, HWE P > 1 * 10^{-5}, call rate > 0.95, and MACH r^2 > 0.5 were used. Details of included cohorts and analytic procedures are described in a previous publication[54].

The GTEx version 8 data included 15,201 samples from 49 tissues of 838 postmortem donors[55]. Meta-analyses of cis-eQTL summary statistics from all 49 tissues were used. Details for samples and analytical procedures are described in another publication[55].

4.4. Reference genome data

Whenever downstream analyses of East Asian summary data required reference genome data, reference panels of East Asian ancestry from the 1000 Genomes Project (1KGP) Phase 3 on GRCh37 or GRCh38 assembly were used, matching with the summary data in each analysis[56]. When assigning genetically inferred ancestries to UK Biobank participants for PWAS weight calculation, a reference panel from the Human Genome Diversity Project (HGDP) was used in addition to the 1KGP panel of all ancestries[57].

4.5. Two-sample, cis-MR

We performed two-sample MR analyses to test causal effects of genetically determined protein levels on the risk of IBD and its subtypes. MR is a method that exploits genetic variants as IV to test a causal hypothesis between an exposure and an outcome[58-60]. In a conventional MR, genetic IVs associated with a modifiable exposure are selected from the entire genomic range. When MR is used for drug target discovery, genetic variants are preferred to be selected from a single genetic region (i.e., cis-variants)[10,61,62]. In an MR using cis-variants (i.e., cis-MR), an exposure that can be pharmacologically perturbed is used, such as a protein level (i.e., pQTL MR) or a gene expression level (i.e., eQTL MR) as in this study.

A cis-MR is subject to the same assumptions in a conventional MR. First, genetic variants must be sufficiently associated with an exposure. Among variants within each genetic region ± 500 kb (i.e., cis-variant), we selected pQTLs with p < 0.05 after adjusting for false discovery rate (FDR) by the Benjamini-Hochberg procedure[63]. Independent IVs were selected after clumping at r^2 > 0.30 within each genetic region[64]. Second, genetic variants must be associated with an outcome only through its association with an exposure (i.e., no horizontal pleiotropy, or exclusion restriction). Restricting IVs to cis-variants offers an advantage of higher specificity in terms of biological mechanisms and has less chance of horizontal pleiotropy, compared to a polygenic approach. In addition, we chose an MR method based on a constrained maximum likelihood (MR-cML), which allows violation of correlated and uncorrelated horizontal pleiotropy and outperformed other methods in a range of simulations[65]. Model averaging and data perturbation were applied to account for model selection uncertainties and potentially large numbers of invalid IVs. Third, genetic variants must not have confounding with an outcome. Confounding by population stratification was controlled by limiting data to East Asian ancestry and utilizing GWAS summary statistics that applied either covariate adjustment of genetic principal components or mixed

effect models. Confounding by LD was assessed with genetic colocalization described below. Transcriptome-wide MR was also conducted with the same procedures as above using eQTLs, instead of pQTLs.

4.6. Colocalization

Colocalization, which assesses whether two traits share common causal variants, can be a useful sensitivity analysis when MR is based on a single gene region[66]. Approximate Bayes factor colocalization (coloc) estimates posterior probabilities (PP) for five hypotheses: H_0, no association with either trait; H_1, association with the first trait but not with the other; H_2, association with the second trait but not with the other; H_3, associations with both traits but separately from distinct variants; H_4, associations with both traits from common variants[67]. High PP-H_4 implies strong evidence that two traits share common causal variants and that confounding by LD is unlikely, whereas high PP-H_3 would suggest that confounding by LD may be present[68]. Some limitations are important to note. Colocalization requires strong statistical genetic associations for both traits, which is hard to achieve when using the East Asian proteome data with a relatively small sample size. Besides, especially with small sample size, differential effects of natural selection on GWAS hits, eQTLs, and pQTLs further complicate colocalization results[69]. Thus, our colocalization results of PP-H_3 and PP-H_4 did not rule out MR findings and were deemed supplementary with a threshold of 0.80. A prior probability of a variant being associated with each trait was set to $1 * 10^{-4}$, and a prior probability being associated with both traits was set to $1 * 10^{-5}$, following a standard practice[68].

4.7. Summary-based PWAS

Summary-based PWAS tests an association between a protein level and an outcome using a protein level imputed from genotypes of an external sample[70,71]. From the UKB individual-level data with protein levels, we first predicted those with East Asian ancestry. Then, we imputed cis-genetic effect on protein, or PWAS weight, using the East Asian samples.

We assigned ancestries to individuals according to a recently proposed method, which is as follows[72]. First, genetic principal component analyses were performed on unrelated individuals from a combined dataset of 1KGP and HGDP. Second, a random forest classifier was trained on using the top six principal components as features and ancestries as labels. Those assigned to East Asian as the final ancestry were used for PWAS weight calculation. We followed details of the ancestry assignment laid out in online tutorials[72].

We used FUSION to compute the PWAS weight[73]. Only genes with statistically significant SNP-based heritability (p < 0.01) were considered. After running multiple predictive models (top pQTL, BLUP, LASSO, Elastic-net) with five cross-validations, the most predictive one was used for calculation.

4.8. Causal-PWAS

Causal-PWAS is a novel extension of PWAS that directly adjust for genetic confounding[74]. PWAS in its original form suffers from confounding by LD; that is, purported cis-genetic effect may in fact be originated from nearby variants or genes, which independently affect a trait of interest. Causal-PWAS jointly models protein effects for all relevant variants and genes simultaneously, rather than for only a selection as in MR or PWAS, and implements a Bayesian variable selection model with sparse prior distributions to identify causal variants or genes. From a causal inference perspective, a distinction is that MR attempts to indirectly overcome confounding with an IV, whereas causal-PWAS directly models potential genetic confounders to estimate conditional causal effects. One important advantage of causal-PWAS is that causal hypothesis testing is allowed in the absence of strong variants (i.e., IVs); MR cannot be performed in this case.

PWAS was first performed as described above. Causal-PWAS takes PWAS Z-score outputs to estimate variant effect and gene effect parameters with sparse prior distributions. Next, using these prior parameters, variant- and gene-level effects are fine-mapped using SuSiE[75,76], which then outputs posterior inclusion probabilities (PIP) for variants and genes. To reduce computational burden, random 10% of variants were used for prior parameter estimation and initial fine-mapping steps. SuSiE was run with an assumption that allows a maximum of five causal variables in a region.

4.9. Multiomics evidence integration

First, by integrating the East Asian GWAS summary statistics for IBD and the East Asian protein data, we identified potential drug targets with MR-cML evidence (FDR-corrected $p < 0.05$) and/or causal-PWAS evidence (PIP > 0.80). We provided MR-cML odds ratios (OR), causal-PWAS PIP, and PWAS Z-scores. Second, for candidate targets, we provided coloc PPs (H_1-H_4), plasma tissue transcriptome-wide MR OR, multi-tissue transcriptome-wide MR OR. All ORs were provided with 95% confidence intervals (CI). We also calculated Pearson correlations between MR-cML estimates of CD and UC to further explore between-subtype similarity in protein effects. All p values in this study were from two-sided tests.

4.10. Gene-level PheWAS database scan

To further evaluate potential pleiotropic or side effects of identified drug targets, we searched through two gene-level phenome-wide association study (PheWAS) databases. The EpiGraphDB is a graph database containing a variety of biomedical and epidemiological relationthips. Its proteome PheWAS browser contains European pQTL MR results for 989 proteins on 225 traits selected from the MR-base database[32]. The AstraZeneca PheWAS Portal lists gene-level rare-variant-aggregated associations for 15,017 traits in the UKB using the exome-wide sequencing data[77,78]. We extracted associations based on thresholds recommended by corresponding databases: $p < 3.5 \times 10^{-7}$ for the EpiGraphDB and $p < 1 \times 10^{-8}$ for the AstraZeneca PheWas.

4.11. Functional enrichment analysis

To explore biological functions of candidate drug targets, enrichment analyses were performed with over-representation tests for Gene Ontology (GO) terms[79], Kyoto Encyclopedia of Genes and Genomes (KEGG)

pathways and modules[80]. GO is a widely used knowledge database that contains computational representations of functions of protein and non-coding RNA molecules, and comprises three orthogonal ontologies: biological process (BP), molecular function (MF), and cellular component (CC). The KEGG pathway map represents a collection of knowledge of the molecular interaction, reaction, and relation networks for metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, human diseases, and drug development. The KEGG modules are manually defined functional units of gene sets, which have a more straightforward interpretation. An over-representation test was used with a default minimum gene set size of 10 to identify GO terms, KEGG pathways, or KEGG modules in which candidate targets are over-represented[81]. Statistical significance was set at $p < 0.05$ after adjusting for FDR by the Benjamini-Hochberg procedure[63].

4.12. Druggable target identification

Druggability of candidate targets was assess based on various drug databases. First, we searched through Open Target Platform[82], ChEMBL[83], DrugBank[84], and Therapeutic Target Database[85] to list drugs approved or in development, which were targeted by candidate targets. By additionally using DailyMed, we extracted indications of approved drugs. Next, with the same process, we identified drugs targeting ligands or receptors of candidate targets. For targets without drugs in development, potential druggability was assessed on the Drug Gene Interaction database (DGIdb)[86].

Declarations

Author contributions

E.K., J.O.K., and S.Y.L. conceptualized and designed the study. E.L., E.K., J.L., Y.J.S. prepared the data. S.Y.L., E.L., and E.K. performed the data analyses. S.Y.L. drafted the manuscript. E.K. and J.O.K revised the manuscript. All authors have read and approved the final version of the manuscript for publication and agreed to be accountable for their contributions.

Conflict of Interest

S.Y.L., E.K., E.L., J.L., Y.J.S., and J.O.K. are employees of Basgenbio, Co., Seoul, Republic of Korea. All authors declare no other competing interests.

Data Availability Statement

Individual data for the UK Biobank can be accessed via request (https://www.ukbiobank.ac.uk/). Summary data are publicly available for IIBDGC IBD GWAS (https://www.ibdgenetics.org/), UKB pQTLs (https://www.synapse.org/#!Synapse:syn51364943/wiki/622119), eQTLGen eQTLs (https://www.eqtlgen.org/phase1.html), GTEx eQTL (https://www.gtexportal.org/home/), 1KGP (https://www.internationalgenome.org/), and HGDP (https://www.internationalgenome.org/).

Code Availability Statement

Data analysis code are available for MR (https://mrcieu.github.io/TwoSampleMR/index.html, https://cran.r-project.org/web/packages/MendelianRandomization/index.html), coloc (https://chr1swallace.github.io/coloc/index.html), ancestry assignment (https://github.com/atgu/hgdp_tgp/tree/master/tutorials), FUSION/PWAS (http://gusevlab.org/projects/fusion/), causal-PWAS (https://xinhe-lab.github.io/ctwas/), functional enrichment analyses (https://github.com/YuLab-SMU/clusterProfiler). Custom code for data preprocessing and analyses are available upon resaonable request.

References

1. Alatab, S. *et al.* The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol* **5**, 17–30 (2020).

2. Shi, H. Y. *et al.* Ethnicity Influences Phenotype and Outcomes in Inflammatory Bowel Disease: A Systematic Review and Meta-analysis of Population-based Studies. *Clinical Gastroenterology and Hepatology* **16**, 190-197.e11 (2018).

3. Yang, Y., Owyang, C. & Wu, G. D. East Meets West: The Increasing Incidence of Inflammatory Bowel Disease in Asia as a Paradigm for Environmental Effects on the Pathogenesis of Immune-Mediated Disease. *Gastroenterology* **151**, e1–e5 (2016).

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. 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).

6. Hong, M. *et al.* Immunochip meta-analysis of inflammatory bowel disease identifies three novel loci and four novel associations in previously reported loci. *J Crohns Colitis* **12**, 730–741 (2018).

7. Cordero, R. Y. *et al.* Trans-ancestry, Bayesian meta-analysis discovers 20 novel risk loci for inflammatory bowel disease in an African American, East Asian and European cohort. *Hum Mol Genet* **32**, 873–882 (2023).

8. Ananthakrishnan, A. N. *et al.* Environmental triggers in IBD: A review of progress and evidence. *Nature Reviews Gastroenterology and Hepatology* vol. 15 39–49 Preprint at https://doi.org/10.1038/nrgastro.2017.136 (2018).

9. Raine, T. & Danese, S. Breaking Through the Therapeutic Ceiling: What Will It Take? *Gastroenterology* **162**, 1507–1511 (2022).

10. Holmes, M. V., Richardson, T. G., Ference, B. A., Davies, N. M. & Davey Smith, G. Integrating genomics with biomarkers and therapeutic targets to invigorate cardiovascular drug development. *Nat Rev Cardiol* **18**, 435–453 (2021).

11. Trajanoska, K. *et al.* From target discovery to clinical drug development with human genetics. *Nature* **620**, 737–745 (2023).

12. Burgess, S. *et al.* Using genetic association data to guide drug discovery and development: Review of methods and applications. *The American Journal of Human Genetics* **110**, 195–214 (2023).

13. Roberts, R. Mendelian Randomization Studies Promise to Shorten the Journey to FDA Approval. *JACC Basic Transl Sci* **3**, 690 (2018).

14. Xu, S. *et al.* Oxidative stress gene expression, DNA methylation, and gut microbiota interaction trigger Crohn's disease: a multi-omics Mendelian randomization study. *BMC Med* **21**, (2023).

15. Mi, J., Wu, X., Bai, X., Yang, Y. & Yang, H. ST2 and CSF-1 as potential druggable targets of inflammatory bowel diseases: Results from two-sample Mendelian randomization study. *Clin Transl Sci* **16**, 236–245 (2023).

16. Chen, J. *et al.* Therapeutic targets for inflammatory bowel disease: proteome-wide Mendelian randomization and colocalization analyses. *EBioMedicine* (2023).

17. Mi, J. *et al.* Mendelian randomization study for the roles of IL-18 and IL-1 receptor antagonist in the development of inflammatory bowel disease. *Int Immunopharmacol* **110**, (2022).

18. Zhuang, Z. *et al.* GWAS-associated bacteria and their metabolites appear to be causally related to the development of inflammatory bowel disease. *Eur J Clin Nutr* **76**, 1024–1030 (2022).

19. Xu, Q. *et al.* Causal Relationship Between Gut Microbiota and Autoimmune Diseases: A Two-Sample Mendelian Randomization Study. *Front Immunol* **12**, (2022).

20. Agrawal, M., Allin, K. H., Petralia, F., Colombel, J. F. & Jess, T. Multiomics to elucidate inflammatory bowel disease risk factors and pathways. *Nature Reviews Gastroenterology and Hepatology* vol. 19 399–409 (2022).

21. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* (2023).

22. Cai, Q., Ding, Z., Fu, A. Z. & Patel, A. A. Racial or ethnic differences on treatment adherence and persistence among patients with inflammatory bowel diseases initiated with biologic therapies. *BMC Gastroenterol* **22**, (2022).

23. Pathiyil, M. M. *et al.* Representation and reporting of diverse groups in randomised controlled trials of pharmacological agents in inflammatory bowel disease: a systematic review. *Lancet Gastroenterol Hepatol* (2023).

24. Nowarski, R. *et al.* Epithelial IL-18 Equilibrium Controls Barrier Function in Colitis. *Cell* **163**, 1444–1456 (2015).

25. Pizarro, T. T. *et al.* IL-18, a Novel Immunoregulatory Cytokine, Is Up-Regulated in Crohn's Disease: Expression and Localization in Intestinal Mucosal Cells. *The Journal of Immunology* **162**, 6829–6835 (1999).

26. Elinav, E. *et al.* NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145**, 745–757 (2011).

27. Wlodek, E. *et al.* A pilot study evaluating GSK1070806 inhibition of interleukin-18 in renal transplant delayed graft function. *PLoS One* **16**, (2021).

28. Hung, L. Y., Pastore, C. F., Douglas, B. & Herbert, D. R. Myeloid-Derived IL-33 Limits the Severity of Dextran Sulfate Sodium–Induced Colitis. *American Journal of Pathology* **191**, 266–273 (2021).

29. Kotsiou, O. S., Gourgoulianis, K. I. & Zarogiannis, S. G. IL-33/ST2 axis in organ fibrosis. *Frontiers in Immunology* vol. 9 (2018).

30. Ngo Thi Phuong, N. *et al.* IL-33 Drives Expansion of Type 2 Innate Lymphoid Cells and Regulatory T Cells and Protects Mice From Severe, Acute Colitis. *Front Immunol* **12**, (2021).

31. Mokry, L. E. *et al.* Interleukin-18 as a drug repositioning opportunity for inflammatory bowel disease: A Mendelian randomization study. *Sci Rep* **9**, (2019).

32. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* **52**, 1122–1131 (2020).

33. Nikolaus, S. *et al.* Evaluation of interleukin-6 and its soluble receptor components sIL-6R and sgp130 as markers of inflammation in inflammatory bowel diseases. *Int J Colorectal Dis* **33**, 927–936 (2018).

34. Mudter, J. *et al.* The transcription factor IFN regulatory factor–4 controls experimental colitis in mice via T cell–derived IL-6. *Journal of Clinical Investigation* (2008).

35. Ito, H. *et al.* A Pilot Randomized Trial of a Human Anti-Interleukin-6 Receptor Monoclonal Antibody in Active Crohn's Disease. *Gastroenterology* **126**, 989–996 (2004).

36. Danese, S. *et al.* Randomised trial and open-label extension study of an anti-interleukin-6 antibody in Crohn's disease (ANDANTE I and II). *Gut* **68**, 40–48 (2019).

37. Zhang, S. *et al.* Effect of Induction Therapy With Olamkicept vs Placebo on Clinical Response in Patients With Active Ulcerative Colitis: A Randomized Clinical Trial. *JAMA* **329**, 725–734 (2023).

38. Krause, P. *et al.* The tumor necrosis factor family member TNFSF14 (LIGHT) is required for resolution of intestinal inflammation in mice. *Gastroenterology* **146**, (2014).

39. Zheng, L., Duan, S. L., Wen, X. L. & Dai, Y. C. Molecular regulation after mucosal injury and regeneration in ulcerative colitis. *Frontiers in Molecular Biosciences* vol. 9 (2022).

40. Takahashi, T. Roles of nAChR and Wnt signaling in intestinal stem cell function and inflammation. *International Immunopharmacology* vol. 81.

41. Acuner-Ozbabacan, E. S. *et al.* The structural network of Interleukin-10 and its implications in inflammation and cancer. *BMC Genomics* **15**, (2014).

42. Bieber, T. Atopic dermatitis: an expanding therapeutic pipeline for a complex disease. *Nature Reviews Drug Discovery* vol. 21 21–40 (2022).

43. Liao, C. *et al.* SPINKs in Tumors: Potential Therapeutic Targets. *Frontiers in Oncology* vol. 12 (2022).

44. Brentnall, T. A. *et al.* Proteins that underlie neoplastic progression of ulcerative colitis. *Proteomics Clin Appl* **3**, 1326–1337 (2009).

45. Csillag, C., Borup, R., Olsen, J., Nielsen, F. C. & Nielsen, O. H. Treatment response and colonic gene expression in patients with Crohn's disease. *Scand J Gastroenterol* **42**, 834–840 (2007).

46. Vieujean, S. *et al.* Potential Role of Epithelial Endoplasmic Reticulum Stress and Anterior Gradient Protein 2 Homologue in Crohn's Disease Fibrosis. *J Crohns Colitis* **15**, 1737–1750 (2021).

47. Karpinska-Leydier, K. *et al.* Correlation Between the Gut Microbiome and Immunotherapy Response in Inflammatory Bowel Disease: A Systematic Review of the Literature. *Cureus* (2021).

48. Dörk, R. *et al.* Myeloid Cell-Specific Deletion of PDGFR-α Promotes Dysbiotic Intestinal Microbiota and thus Increased Colitis Susceptibility. *J Crohns Colitis* (2023).

49. Jacob, J. M. *et al.* PDGFRα-induced stromal maturation is required to restrain postnatal intestinal epithelial stemness and promote defense mechanisms. *Cell Stem Cell* **29**, 856-868.e5 (2022).

50. Li, H. *et al.* Nintedanib Alleviates Experimental Colitis by Inhibiting CEBPB/PCK1 and CEBPB/EFNA1 Pathways. *Front Pharmacol* **13**, (2022).

51. Eldjarn, G. H. *et al.* Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348–358 (2023).

52. Sun, B. B. *et al.* Genetic regulation of the human plasma proteome in 54,306 UK Biobank participants 2 3. (2023).

53. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097–1103 (2021).

54. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* **53**, 1300–1310 (2021).

55. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues The GTEx Consortium*. *Science (1979)* 1318–1330 (2020).

56. Auton, A. *et al.* A global reference for human genetic variation. *Nature* vol. 526 68–74 (2015).

57. Bergström, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science (1979)* **367**, (2020).

58. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Great Britain International Journal of Epidemiology* **32**, 1–22 (2003).

59. Sanderson, E. *et al.* Mendelian randomization. *Nature Reviews Methods Primers* **2**, 1–21 (2022).

60. Richmond, R. C. & Smith, G. D. Mendelian Randomization: Concepts and Scope. *Cold Spring Harb Perspect Med* **12**, a040501 (2022).

61. Gill, D. *et al.* Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Res* **6**, 24 (2021).

62. Daghlas, I. & Gill, D. Mendelian randomization as a tool to inform drug development using human genetics. *Cambridge Prisms: Precision Medicine* **1**, e16 (2023).

63. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300 (1995).

64. Gkatzionis, A., Burgess, S. & Newcombe, P. J. Statistical methods for cis-Mendelian randomization with two-sample summary-level data. *Genet Epidemiol* (2022).

65. Xue, H., Shen, X. & Pan, W. Constrained maximum likelihood-based Mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *Am J Hum Genet* **108**, 1251–1269 (2021).

66. Burgess, S. *et al.* Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res* (2023).

67. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet* **10**, (2014).

68. Zuber, V. *et al.* Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. *The American Journal of Human Genetics* (2022).

69. Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat Genet* (2023).

70. Wingo, T. S. *et al.* Brain proteome-wide association study implicates novel proteins in depression pathogenesis. *Nat Neurosci* **24**, 810–817 (2021).

71. Wingo, A. P. *et al.* Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nat Genet* **53**, 143–146 (2021).

72. Koenig, Z. *et al.* A harmonized public resource of deeply sequenced diverse human genomes. (2023).

73. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* **48**, 245–252 (2016).

74. Zhao, S. *et al.* Adjusting for genetic confounders in transcriptome-wide association studies leads to reliable detection of causal genes. (2022).

75. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the 'Sum of Single Effects' model. *PLoS Genet* **18**, (2022).

76. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Series B Stat Methodol* **82**, 1273–1300 (2020).

77. Dhindsa, R. S. *et al.* Rare variant associations with plasma protein levels in the UK Biobank. *Nature* (2023).

78. Wang, Q. *et al.* Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* **597**, 527–532 (2021).

79. Ashburner, M. *et al. Gene Ontology: tool for the unification of biology The Gene Ontology Consortium**. (2000).

80. Kanehisa, M. & Goto, S. *KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research* vol. 28 (2000).

81. Boyle, E. I. *et al.* GO::TermFinder - Open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20**, 3710–3715 (2004).

82. Ochoa, D. *et al.* The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res* **51**, D1353–D1359 (2023).

83. Mendez, D. *et al.* ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res* **47**, D930–D940 (2019).

84. Wishart, D. S. *et al.* DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res* **46**, D1074–D1082 (2018).

85. Zhou, Y. *et al.* TTD: Therapeutic Target Database describing target druggability information . *Nucleic Acids Res* (2023).

86. Freshour, S. L. *et al.* Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res* **49**, D1144–D1151 (2021).
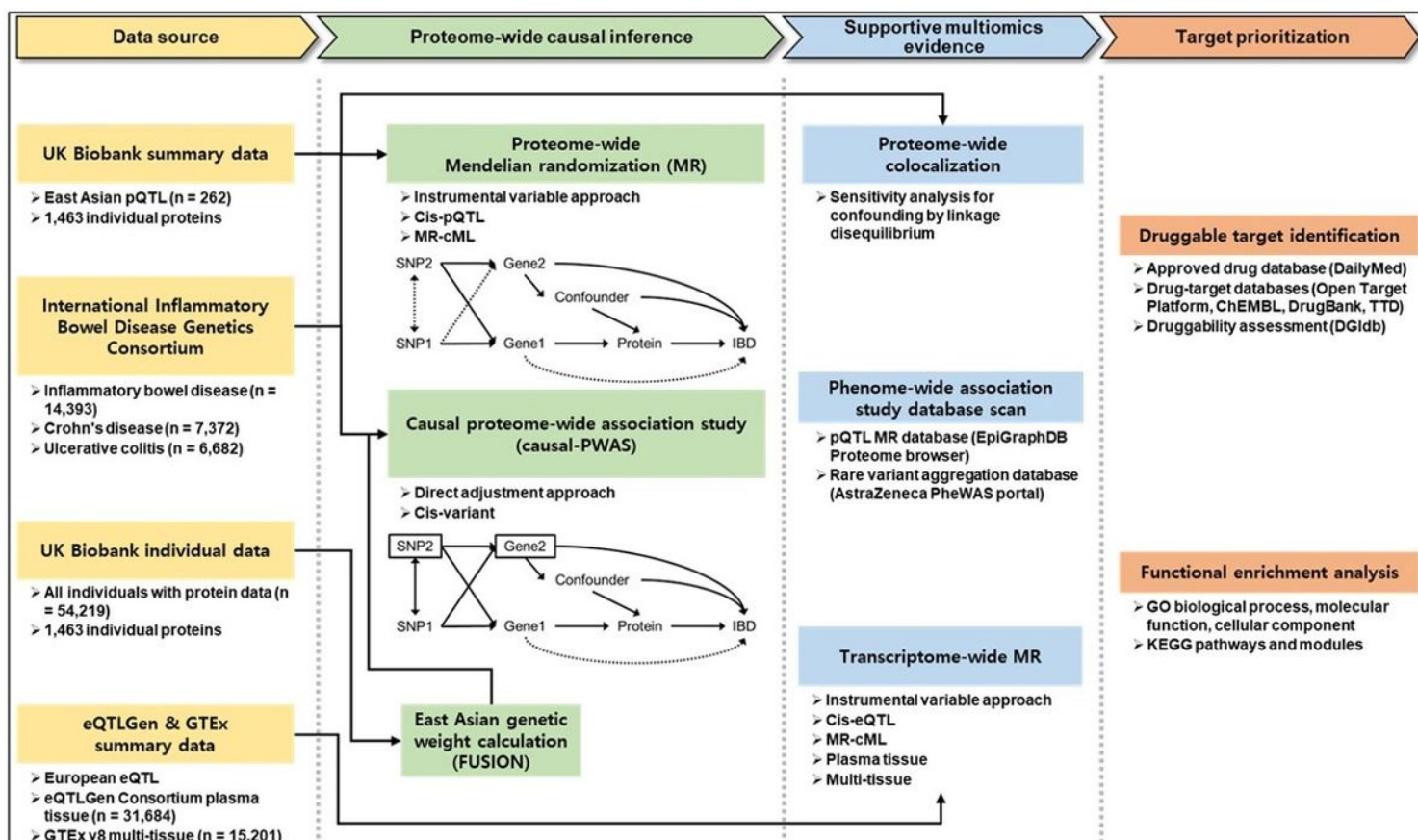
Figures
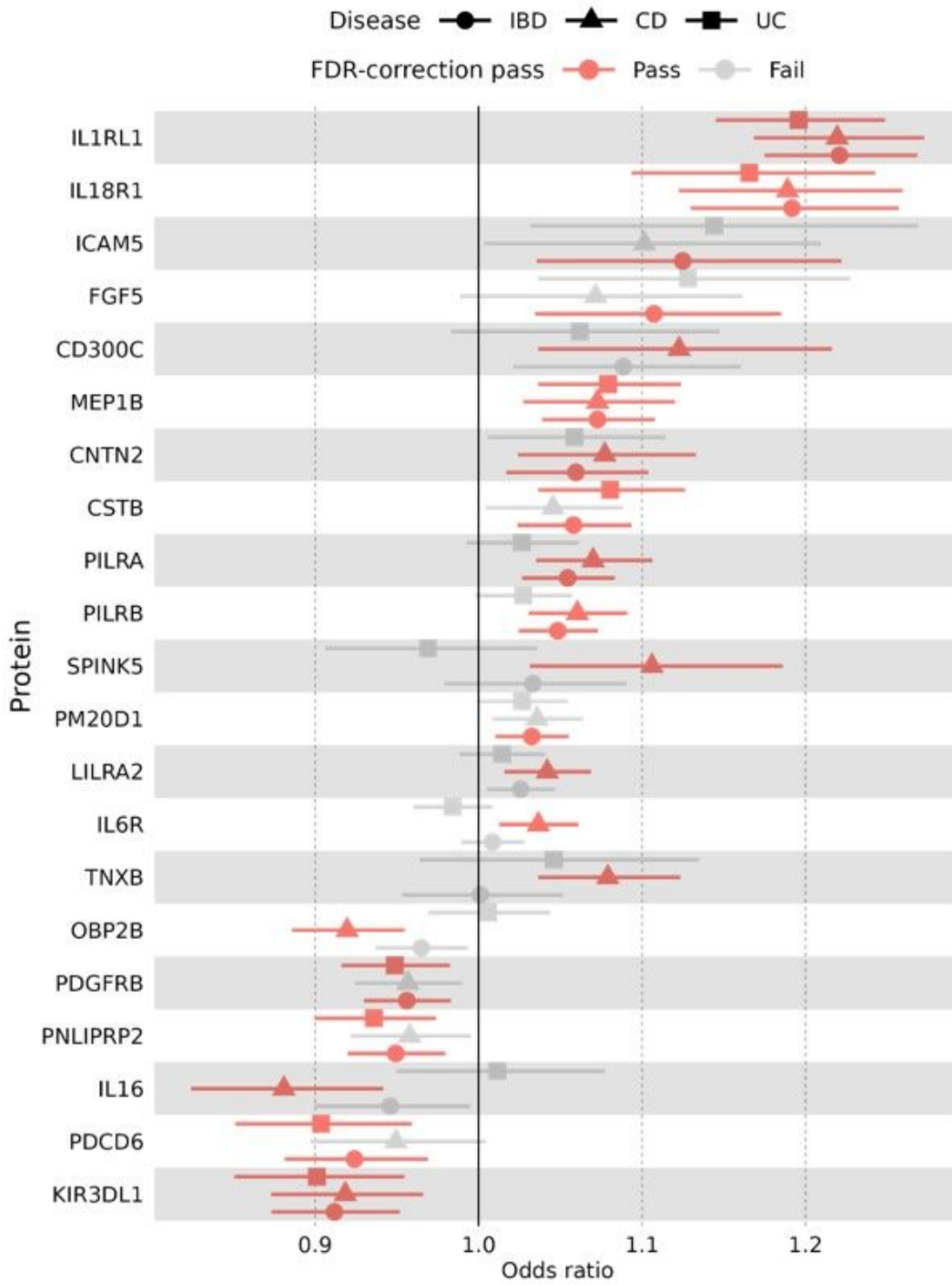
Figure 1

An overview of the study.

Figure 2

Proteome-wide MR results for IBD and its subtypes. MR results of proteins with at least one statistically significant association are presented. Circles, trigangles, and squares specify IBD, CD, and UC, respectively. Dots denote odds ratios, and lines denote unadjusted 95% confidence intervals. Lines are red if the association passed the FDR correction threshold.

IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; FDR, false discovery rate.
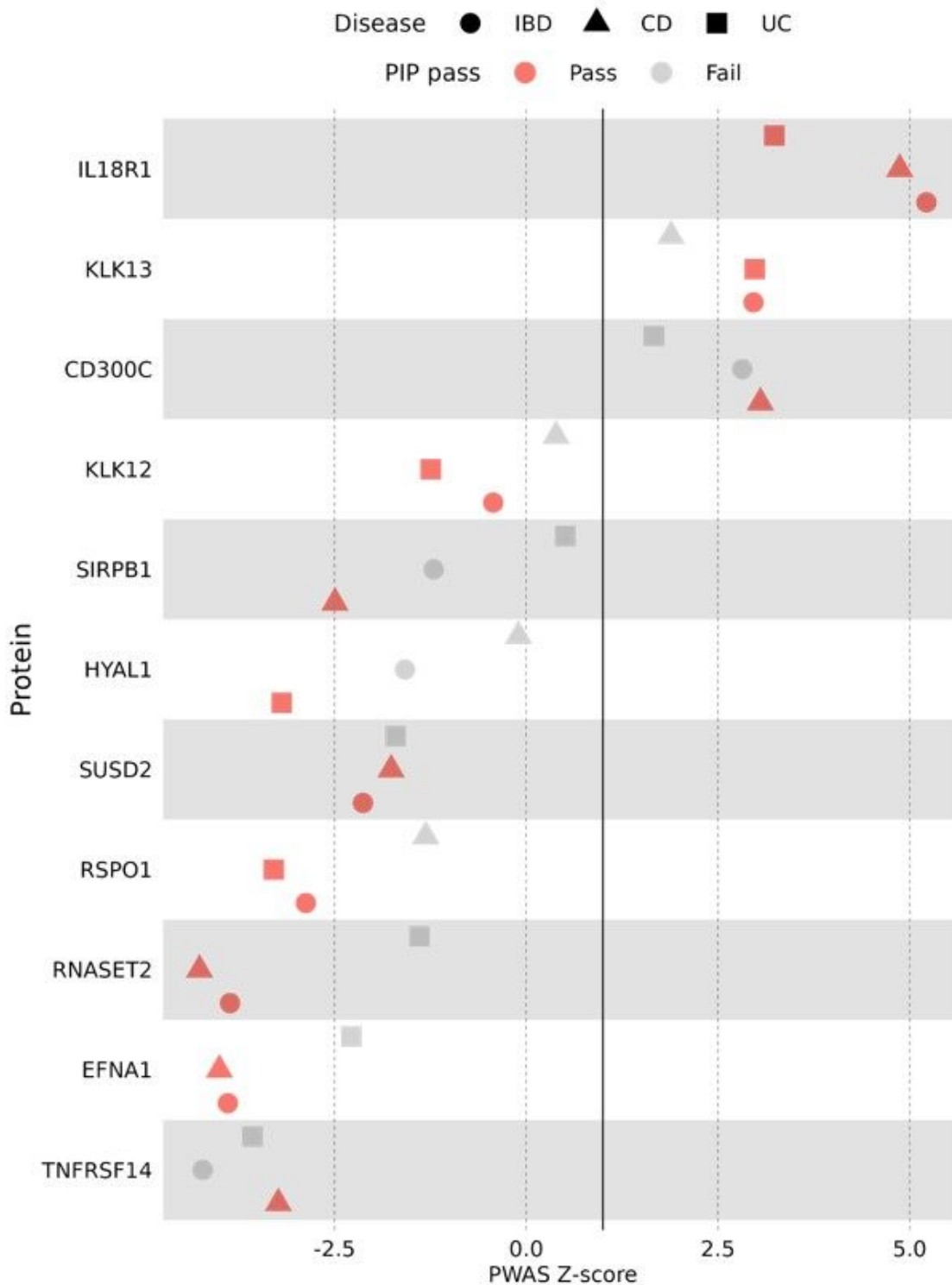
Figure 3

PWAS and causal-PWAS extension results for IBD and its subtypes. PWAS results for proteins with at least one association that passed the causal-PWAS PIP threshold. Circles, triangles, and squares specify IBD, CD, and UC, respectively. Dots denote PWAS Z-scores and are red if the association passed the 80% PIP threshold.

IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; PIP, posterior inclusion probability; PWAS, proteome-wide association study.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalMaterials.docx