Bi-directional Mendelian randomization analysis provides evidence for the causal involvement of dysregulation of CXCL9, CCL11 and CASP8 in the pathogenesis of ulcerative colitis

Bi-directional Mendelian randomization analysis provides evidence for the causal involvement of dysregulation of CXCL9, CCL11 and CASP8 in the pathogenesis of ulcerative colitis

Running title: Systemic inflammation and ulcerative colitis

Authors: Jie Chen[1,2,3†], Yajing Zhou[1†], Yuhao Sun[2†], Shuai Yuan[1,4], Rahul Kalla[5], Jing Sun[1], Jianhui Zhao[1], Lijuan Wang[1], Xuejie Chen[3], Xuan Zhou[1], Siqi Dai[6], Yu Zhang[7], Gwo-tzer Ho[5], Dajing Xia[8], Qian Cao[7], Zhanju Liu[9], Susanna C. Larsson[4,10], Xiaoyan Wang[3], Kefeng Ding[6#], Jonas Halfvarson[11], Xue Li[1#], Evropi Theodoratou[12,13], Jack Satsangi[14]

[1] Department of Big Data in Health Science, School of Public Health and The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

[2] Centre for Global Health, Zhejiang University School of Medicine, Hangzhou, China

[3] Department of Gastroenterology, The Third Xiangya Hospital, Central South University, Changsha, China

[4] Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

[5] Edinburgh IBD Science Unit, Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK

[6] Colorectal Surgery and Oncology, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Zhejiang, China

[7] Department of Gastroenterology, Sir Run Run Shaw Hospital, College of Medicine Zhejiang University, Hangzhou, China

[8] Department of Toxicology of School of Public Health, & Center of Immunology & Infection, Zhejiang University School of Medicine, Hangzhou, China

[9] Center for IBD Research, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China

[10] Unit of Medical Epidemiology, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

[11] Department of Gastroenterology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

[12]Centre for Global Health, Usher Institute, University of Edinburgh, Edinburgh, UK

[13] Cancer Research UK Edinburgh Centre, Medical Research Council Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK

[14] Translational Gastroenterology Unit, Nuffield Department of Medicine, Experimental Medicine Division, University of Oxford, John Radcliffe Hospital, Oxford, UK

† Chen J, Zhou Y, and Sun Y shared first authorship.* Satsangi J, Theodoratou E and Li X share last authorships.

 # Corresponding authors:

Xue Li, School of Public Health and the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China; E-mail: xueli157@zju.edu.cn; Kefeng Ding, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China; E-mail: kefengding@zju.edu.cn;

Authors contributions:

Study conceptualization and design: X.L., J.S., and E.T.; Data collection and curation: J.C., Y.H.S., X.L., S.C.L., E.T. and J.S.; Methodology establishment: J.C., Y.J.Z., S.Y., L.J.W., X.Z. and X.L.; Formal analysis and software support: Y.J.Z., S.Y., J.C., and X.L.. The original draft was written by Y.J.Z., S.Y., J.C., and X.L.; and the manuscript was reviewed and edited by X.L., E.T., J.S., S.Y., R.K. and Z.J.L.. All co-authors had the opportunity to comment on the analysis and interpretation of the findings and approved the final version for publication.

Abbreviations:

UC, ulcerative colitis;

TSMR, two-sample Mendelian randomization;

GWAS, genome-wide association study;

pQTLs, protein quantitative trait loci;

CXCL, C-X-C motif chemokine ligand;

CCL, C-C motif chemokine ligand;

IBD, inflammatory bowel disease;

IL, interleukin;

TNF, tumor necrosis factor;

NPX, normalized protein expression;

LOD, limit of detection;

CI, confidence interval;

SNP, single nucleotide polymorphism;

LD, linkage disequilibrium;

IV, instrumental variable;

FDR, false discovery rate;

TGF, transforming growth factor;

FC, fold change;

OSM, oncostatin M;

HGF, hepatocyte growth factor;

MMP10, matrix metallopeptidase 10;

ENRAGE, extracellular newly identified RAGE-binding protein;

MCP, monocyte chemoattractant protein;

IFN, interferon;

DSS, deficient in sulfate sodium;

AA, acetic acid;

RIP, release of interacting protein;

Abstract

Background and Aims

Systemic inflammation is well-recognized to be associated with ulcerative colitis (UC), but whether these effects are causal or consequential remains unclear. We aimed to define potential causal relationship of cytokine dysregulation with different tiers of evidence.

Methods

We firstly synthesized serum proteomic profiling data from two multi-centered observational studies, in which a panel of systemic inflammatory proteins was analyzed to examine their associations with UC risk. To further dissect observed associations, we then performed a bidirectional two-sample Mendelian randomization (TSMR) analysis from both forward and reverse directions using five genome-wide association study (GWAS) summary level data for serum proteomic profiles and the largest GWAS of 28,738 European-ancestry individuals for UC risk.

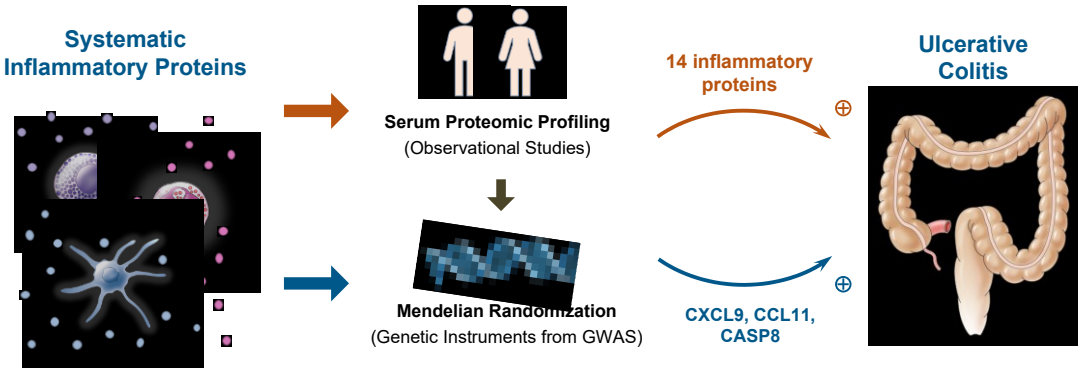
Results

Pooled analysis of serum proteomic data identified 14 proteins to be associated with the risk of UC. Forward MR analysis using only cis-acting protein quantitative trait loci (cis-pQTLs) or trans-pQTLs further validated causal associations of two chemokines and the increased risk of UC: C-X-C motif chemokine ligand 9 (CXCL9) (OR, 1.45, 95% CI, 1.08-1.95, $P=.012$) and C-C motif chemokine ligand 11 (CCL11) (OR, 1.14, 95%CI: 1.09-1.18, $P=3.89\times10^{-10}$). Using both cis- and trans-acting pQTLs, an association of caspase-8 (CASP8) (OR, 1.04, 95% CI, 1.03-1.05, $P= 7.63\times10^{-19}$) was additionally identified. Reverse MR did not find any influence of genetic predisposition to UC on any of these three inflammation proteins.

Conclusions

Pre-existing elevated levels of CXCL9, CCL11 and CASP8 may play a role in the pathogenesis of UC.

Keyword: systemic inflammatory proteins; ulcerative colitis; Mendelian randomization;



104

105 **Graphical abstract**

Introduction

Ulcerative colitis (UC) is a subtype of inflammatory bowel disease (IBD) characterized by chronic relapsing inflammation of the colon, on the background of a systemic immune response[1]. The incidence of UC has been increasing worldwide and the prevalence rate is projected to be 1% among Western populations by 2030[2], posing a substantial burden on global health[3,4].

Although the cause of UC remains unclear, systemic inflammation is a hallmark feature of this disease, triggered by immune dysregulation in response to the exposome in a genetically susceptible individual[5,6]. A wide range of inflammatory cytokines, including interleukin-1 receptor antagonist (IL1R), IL12, IL23, IL17, tumor necrosis factor (TNF) alpha, C-C motif chemokine ligand (CCL) 2 and 3 have been reported to be associated with UC risk in observational studies[7]. These findings underline the complexity of intestinal inflammation in UC and indicate the possibility of cytokines as potential targets for primary prevention and therapy of UC. Current evidence supports that TNF-mediated as well as several non-TNF pathways (e.g., IL-1, IL-12/23) are involved in disease pathogenesis and represent targets for current and future therapeutic strategies of UC[8,9]. However, the causal role of specific inflammatory pathways remains unclear, primarily driven by the limitations of observational studies (e.g., residual confounding and reverse causality) and the lack of high-quality data from randomized trials. These research gaps and limitations preclude the differentiation between the *causal* role of pre-existing systemic inflammation in UC onset and the *consequential* effect of UC progression on systemic inflammation. This distinction has important clinical implications when exploring novel therapeutic targets or identifying predictors of drugresponse or prognostic markers of future disease course for UC.

By employing genetic variants as instrumental variables for an exposure (e.g., levels of a specific cytokine), Mendelian randomization (MR) analysis can strengthen causal inference by minimizing unobserved confounding and diminishing reverse causality[10]. Because genetic variants are randomly allocated at conception, under certain assumptions, the MR framework mimics randomized controlled trials and enables more robust causal inference than traditional epidemiologic techniques. Here, we conducted observational and MR analysis to provide different

tiers of evidence to elaborate the exact role of proteomic signatures of systemic inflammation in the onset of UC.

Methods

Study design

In this study (**Figure 1**), we firstly synthesized serum proteomic profiling data from two published multi-center observational studies[11,12] to examine the associations between systemic inflammatory proteins and the risk of UC. The case-control study was adopted as discovery dataset and the prospective case-cohort study as validation dataset. Next, bidirectional two-sample MR was then performed to comprehensively examine the associations between circulating inflammatory proteins and UC risk from both forward and reverse directions.

Two-stage and pooled analyses of serum proteomic data from multi-centered observational studies

Study populations

The first discovery case-control study included 159 UC patients and 293 non-IBD controls who were recruited across six clinical centers in Europe[11]. Blood samples for protein profiles were collected at baseline, i.e., at the time of recruitment. The proteins measured in this study were involved in various UC-related mechanisms, including inflammation, immune regulation, metabolism, and cell-cell signaling. The second prospective nested case-cohort study for validation included 72 preclinical UC cases and 140 matched healthy controls who were recruited by the Northern Sweden Health and Disease Study in Sweden (NSHDSS)[12]. Participants who developed UC later in life were identified by linking the Northern Sweden Health and Disease Study register dataset with the International Classification of Diseases code register of Region of Västerbotten, Sweden. Pre-diagnostic plasma samples from UC patients and matched controls who remained free from IBD during follow-up were measured for 92 proteins related to inflammation.

Quality control

To minimize inter- and between-study variations, the pre-processed proteomic data were unified as arbitrary units, i.e., normalized protein expression [NPX] on a $\log_2$ scale. A high NPX represents

high protein concentration and a low NPX represents low protein concentration. The limit of detection [LOD] for each protein probe was defined as the mean plus three standard deviations of the negative controls. The proteins for which >50% of samples were below the LOD were excluded from the analysis. After quality control and compiling the proteins together, a total of 57 common proteins related to systemic inflammation were eligible to be analyzed in 231 UC patients and 433 controls.

Statistical analysis

We used multivariable logistic regression models, with age, sex, smoking status and Montreal classification as covariates, to investigate associations between individual proteins and risk of UC in discovery and validation study populations separately. A pooled analysis of individual data from the discovery and validation studies was furtherly conducted to compute a pooled effect estimates with its 95% confidence interval (CI) with adjustment of age, sex, smoking status, Montreal classification and study as covariates, for the associations between inflammation proteins and the risk of UC.

Causal inference —Bidirectional two-sample MR analysis

Data source for inflammatory proteins

MR analysis was conducted for the 14 inflammation proteins which were implicated by replication of association in both the discovery and replication datasets. In the forward MR, single nucleotide polymorphisms (SNPs) strongly associated with systemic inflammation protein levels ($P < 5\times10^{-8}$) (known as protein quantitative trait loci, pQTLs) were obtained from a published Phenome-wide Mendelian randomization of plasma proteosome[13] and five genome-wide association studies (GWASs)[14-18] (**Supplementary Table 1**). In the reverse MR, genetic associations were derived from the most recent and largest GWAS of plasma proteins measured with 4,907 aptamers of 35,559 Icelanders[17]. Linkage disequilibrium (LD) was calculated based on 1000 Genomes European reference panel, and genetic variants without LD ($r^2 \leq 0.001$ and clump window >10,000 kb) were finally selected as independent instrumental variables (IVs). The selected genetic IVs

were further classified as: either 1) cis-pQTLs located in the vicinity of the encoding gene (defined as ≤500 kb from the leading pQTL of the test protein); or 2) trans-pQTLs located outside this window. MR analyses separately using cis- and trans-acting pQTLs and combined overall analyses were performed to test the robustness of the MR findings. Detailed information on the SNPs used as genetic IVs for inflammation proteins is presented in **Supplementary Table 2**.

Data source for ulcerative colitis

Genetic associations between the cytokines-related SNPs and UC were obtained from the largest GWAS summary data of the international IBD genetic consortium (IIBDGC) study, which included 6,968 UC cases and 21,770 population controls of European descent. Genetic variants associated with UC at the genome-wide significance level ($P < 5\times10^{-8}$). After pruning SNPs in linkage disequilibrium ($r^2 \leq 0.001$ and clump window >10,000 kb), 42 SNPs were selected as genetic IVs of UC to be used in reverse MR analyses (**Supplementary Table 3**).

Two-sample MR analysis

For both the forward (the effect of circulating cytokines on UC) and reverse (the effect of genetic liability to UC on circulating cytokine levels) two-sample MR analyses, the inverse-variance weighted method with random-effects was used as the main method. Four sensitivity analyses were conducted for supplement, including the weighted median[10], MR-Egger regression[19], MR-PRESSO[20] and leave-one-out analysis. Wald ratio was calculated for every single SNP to estimate the association between exposure and outcome. The weighted median method can provide consistent causal estimates if more than 50% of the weight comes from valid instrumental variables[10]. MR-Egger can generate estimates after correcting for horizontal pleiotropy; however, this method compromises statistical power[19]. MR-PRESSO can detect outlying instrumental variables and provide causal estimates after removal of these outliers[20]. Cochrane's Q value was used to assess the heterogeneity among estimates of genetic instruments and the *p* value for intercept in MR-Egger was used to detect horizontal pleiotropy[19]. The strength of instruments was assessed by calculating F-statistics (F<10 was deemed as a weak instrument)[21]. The threshold of

212 statistical significance was corrected by false discovery rate (FDR) for multiple comparison: FDR
213 <0.05 was regarded significant and p <0.05 but not survived FDR was regarded suggestive
214 significant. All statistical analyses were two-sided and performed in R 4.0.3 software using the R
215 packages 'TwoSampleMR'[22] and 'MR-PRESSO'[20].

216 All studies were approved by the irrespective institutional review boards and conducted with
217 appropriate ethical criteria in each country and in accordance with the Declaration of Helsinki.

Results

The basic characteristics of the study populations are presented in **Table 1**. In the discovery cohort, the mean age of participants was 37.3 vs 32.4 years, and the proportion of males was 57.9% vs 45.4% in UC patients and controls. In the validation cohort, the mean age was 48.3 vs 48.1 years, and the proportion of males was 47.2% vs 45.7% in UC patients and controls. The detailed results of individual and pooled analyses of serum proteomic profiling data for 57 systemic inflammatory protein markers from discovery and validation datasets were summarized in **Supplementary Table 4**.

In the discovery dataset, 11 systemic inflammatory proteins were found to be significantly higher in UC patients than controls after correction of multiple testing by FDR, including transforming growth factor alpha (TGFA) (log_2fold change [log_2FC]=0.20, $P=1.16\times10^{-4}$), oncostatin M (OSM) (log_2FC=0.10, *P=.003*), hepatocyte growth factor (HGF) (log_2FC=0.07, *P=.015*), matrix metallopeptidase 10 (MMP10) (log_2FC=0.10, *P=.015*), TNF superfamily member 14 (TNFSF14) (log_2FC=0.06, *P=.016*), IL6 (log_2FC=0.15, *P=.021*), C-X-C motif chemokine ligand (CXCL) 1 (log_2FC=0.06, *P=.015*), CCL20 (log_2FC=0.06, *P=.024*), extracellular newly identified RAGE-binding protein (EN-RAGE) (log_2FC=0.07, *P=.025*), CXCL9 (log_2FC=0.11, *P=.030*), and CCL4 (log2FC=0.04, *P=.032*). **(Figure 2)** In the validation dataset, 2 out of 11 protein markers were successfully replicated in a prospective study design: MMP10 (log_2FC=0.07, *P=.015*) and CXCL9 (log_2FC=0.06, *P=.015*), and three more cytokines, including CXCL11 (log_2FC=0.03, *P=.035*), CCL11 (log_2FC=0.03, *P=.016*) and monocyte chemoattractant protein-1 (MCP1) (log2FC=0.02, P=*.037*) were additionally reported indicating that individuals who developed UC during the follow-up period had significantly higher serum levels of these inflammatory markers at baseline **(Figure 2)**.

In the pooled analyses of the two datasets, 14 out of 57 systemic inflammatory makers showed significant associations with UC risk after correction of multiple testing by FDR, including TGFA (log_2FC=0.17, $P=1.16\times10^{-4}$), MMP10 (log_2FC=0.08, *P=.001*), OSM (log_2FC=0.10, *P=.010*), CXCL1 (log_2FC=0.04, *P=.021*), CXCL9 (log_2FC=0.09, *P=.010*), HGF (log_2FC=0.04, *P=.016*), TNFSF14

(log$_2$FC=0.06, *P=.019*), IL6 (log$_2$FC=0.14, *P=.021*), ENRAGE (log2FC=0.081, *P=.021*), CCL4 (log$_2$FC=0.03, *P=.024*), CCL20 (log$_2$FC=0.06, *P=.029*), CASP8 (log$_2$FC=0.114, *P=.037*), CCL11 (log$_2$FC=0.02, P=*.037*), and IL8 (log$_2$FC=0.05, *P=.040*) **(Figure 2)**. The robustness of these associations was strengthened in sensitivity analyses: 1) stratified analysis of smoking status separate in two cohorts and combined, with adjustments of age, sex and Montreal classifications and study effect **(Figure S1-3)**; 2) stratified analysis of Montreal classifications and study in validation cohort, with adjustments of age, sex, and smoking status **(Figure S4)**.

In both forward and reverse MR, the F-statistics for used genetic instruments were all over 10, suggesting no substantial weak instrument bias (**Supplementary Table 2, 3**). Detailed information and results of the forward MR analysis are shown in **Figure 3** and summarized in **Supplementary Table 5**. Forward MR analyses were performed for 13 out of 14 inflammatory proteins identified from the observational study, for which there were available genetic IVs (we found no genetic eligible genetic IV for TGFA). CXCL9 (OR, 1.45, 95% CI, 1.08-1.95; *P=.012;* per SD increment) and CCL11 (OR, 1.14, 95%CI: 1.09-1.18, $P=3.89\times10^{-10}$; per SD increment) were suggested to have causal effect on the UC risk respectively using only cis- or trans-acting pQTLs. CASP8 (OR, 1.04, 95% CI, 1.03-1.05, $P=7.63\times10^{-19}$; per SD increment) were additionally found to have significant associations with increased risk of UC using both cis- and trans-acting pQTLs. In the reverse MR analysis, a total of 42 SNPs strongly associated with UC were included as genetic instruments (**Supplementary Table 3**). Results from IVW MR analysis showed that genetic liability to UC was associated with none of the three inflammatory proteins (CXCL9, CCL11, CASP8) reported in forward MR. No indications of horizontal pleiotropy were detected by MR-Egger intercept test, and two outliers were found with CCL11 by the outlier test of MR-PRESSO analyses, but there were no significant differences before and after the correction. The results of reverse MR analysis are presented in **Table 2**.

Discussion

The potential causal role of systemic inflammation in the pathogenesis of UC has not been fully established. A major advantage, laying the foundation of both novelty and rigor of this study, is the adoption of the MR approaches together with the profiling of serum proteome to provide different tiers of evidence to cross-validate the study findings. We compiled observational datasets and performed a pooled analysis of serum proteomic profiling data to derive primary clues for observational associations between inflammatory proteins and UC risk. The bidirectional MR study was then conducted based on inflammatory proteins firstly screened out, contributing to estimate both directions of association between inflammatory proteins and UC. These findings therefore provides us with major insights into a more precise causal inference about the exact role of cytokines in the onset or progress of UC. Five GWASs of systemic inflammatory proteins and one PheWAS-MR study of plasma proteosome were employed as our data sources, comprehensively enlarging our investigation scope. Specifically, the pooled analysis of serum proteomic profiling data suggested that 14 systemic inflammatory proteins (TGFA, MMP10, OSM, CXCL1, CXCL9, HGF, TNFSF14, IL6, IL8, CCL4, CCL11, CCL20, CASP8 and ENRAGE) were significantly higher in preclinical or newly diagnosed UC patients than controls. The forward MR analysis using cis-pQTLs and trans-pQTLs indicated that genetically predicted high circulating levels of CXCL9, CCL11 and CASP8 were associated with increased risk of UC, and the reverse MR analysis did not find any influence of genetic predisposition to UC on the circulating levels of these inflammatory proteins.

Proteomic data employed in our analysis were derived from a case-control study (discovery cohort) and a prospective nested case-cohort study (replication cohort), providing us with sufficient and enhanced strength of evidence. The discovery cohort recruited participants across six centres in Europe to create and evaluate novel multi-protein panels based on known or suspected involvement in the pathogenesis of UC. The replication cohort was specially focused on UC developed later in life, which aimed to investigate the association between inflammatory proteins and future diagnosis of UC. This cohort of preclinical ulcerative colitis comprehensively measured

297 the effects of 92 predefined inflammatory proteins and was additionally employed to derive
298 biosignature models and further validate the associations with a broader basis of the population.
299 Given the inherent limitations of observational design and the fact that inflammation is well known
300 to accompany diagnosis and often resolves quickly, these observed associations could only be
301 implied as correlation rather than causation. We therefore took advantage of genetic instruments
302 to proxy the long-term exposure to systemic inflammation and adopted a MR design to further
303 explore any causal relationships.

304 High serum levels of CXCL9 were found to be causally associated with increased risk UC in the
305 pooled analysis of serum proteomic profiling data and forward MR analysis utilizing pQTLs. CXCL9
306 is a chemokine of C-X-C subfamily[23] expressed in macrophages under activation of interferon-
307 gamma (IFN)-α[24] and IFN-γ[25]. Numerous studies have reported that CXCL9 levels are increased
308 in several autoimmune diseases such as type 1 diabetes[26,27] and systemic rheumatological
309 disorders[28-30], indicating a potential association of CXCL9 with inflammatory diseases. A
310 therapeutic strategy targeting the CXCL9 pathway has been indicated in a previous study,[31] where
311 curcumin phytochemicals were reported to inhibit the CXCL9 inflammatory cascade. Collectively,
312 these findings underscore the importance of CXCL9 in disease onset and may pave a new way for
313 the prevention and treatment of UC.

314 Serum CCL11 levels were also indicated to have a strong association with UC risk in the pooled
315 analysis of proteomic data, and this observational association was further validated in MR analysis
316 by utilizing pQTLs. CCL11, also known as Eotaxin-1, is another member of C-C motif chemokine
317 subfamilies[32] and can act as a chemoattractant of eosinophils and other leukocytes in inflammatory
318 diseases.[33] A previous case-control study conducted in 35 patients with UC and 38 healthy controls
319 reported significantly higher serum levels of Eotaxin (-1) in patients with UC, providing early
320 observational evidence for the contribution of Eotaxin to the pathogenesis of UC.[34] Another nested
321 case-control study including 137 UC cases and 38 healthy controls, found the levels of Eotaxin-1
322 to be elevated in patients with histologically active UC vs controls using high-throughput
323 technologies (Luminex-based multiplex testing) of 42 analytes, suggesting Eotaxin-1 to be an

essential etiologic factor and a potential treatment target.[35] Based on the effect of CCL11 mediated by activation of inflammatory cytokines and induction of tissue damage on the exacerbation of UC, some experimental studies have put an effort in investigating the CCL11-related path for treatment utilizing monoclonal antibody (anti-eotaxin-1) or constructing CCL11 deficient in sulfate sodium (DSS)-induced colitis mice,[36,37] further demonstrating the potentially value in the development of CCL11-targeted therapies.

Our study also reported that elevated levels of CASP8 were causally associated with increased risk of UC in both the pooled analysis of proteomic data and the forward MR analysis. CASP8 is a cysteine protease capable of inducing a process called “extrinsic apoptosis” via death receptors (members of TNF subfamily)[38], and it can also exert other non-apoptosis effects like regulating embryogenesis and cell proliferation[39]. A recent cross-sectional study including 40 patients with active UC and 21 healthy controls suggested that CASP8 was a distinguishing protein in the systemic inflammatory protein profile showing an increased level[40], which possibly takes effect via interacting with inflammatory factors[41]. Also, some studies provided supportive experimental evidence that the expression levels of CASP8 were significantly elevated in the colonic mucosa of rats with acetic acid (AA)-induced UC,[42] which could be suppressed via TLR4/NF-κB-related pathway (e.g. betulin attenuates), TGF-β-related signaling (e.g. Cyclosporine)[43], etc., revealing a ponderable perspective on the development of CASP8-targeted drugs or comprehensive treatments. Nonetheless, some studies indicated opposite findings that CASP8 was reduced in UC patients[44], and the deficiency of CASP8 could up-regulate the inflammasome activity via increasing the release of interacting protein 3 (RIP3)[45]. Given that experimental studies on animals are insufficient to achieve a broad extrapolation of conclusions, further practice of CASP8 in human population are worthy carrying out to confirm its pathological mechanisms in UC.

Numerous studies have proposed that lifestyle factors, such as dietary intakes, stress and physical activities have impacts on the onset and progress of UC, and indicated that a healthy lifestyle is indicated to be an effective strategy to reduce the systemic inflammation in UC.[46] Diet with sufficient intake of fruits and vegetables may reduce the systemic inflammation[47] , and ong-term aerobic

exercising was also reported to reduce serum levels of proinflammatory cytokines in older adults[48]. Both of these findings suggested the role of healthy diet and physical activity in alleviating inflammatory diseases like UC. With the establishing role of these cytokines/chemokines in UC progress, efforts towards formulating pharmacological (potential drugs or phytochemicals) or non-pharmacological strategies (anti-inflammatory diet, exercise, or stress reduction) via the regulation of CXCL9, CCL11 and CASP8 should be considered as possible strategies for primary prevention of UC.

Limitations

Notwithstanding the strengths of our study, there are also some potential limitations. In this study, we only conducted MR analysis with 13 inflammatory proteins that had been identified by a pooled analysis of two serum proteomic datasets, which constrained the range of cytokines assessed such that we were not able to verify potentially effective cytokines reported in our analyses of observational studies (e.g. TGFA) or suggested by previous studies, like IL23, IL7, etc. Another limitation of our study is that utilizing SNPs as instruments to deduce causality in MR analyses may risk horizontal pleiotropy, especially for protein markers proxied by a few SNPs. In addition, the confinement of population to European descent is important in considering the generalizability of our findings to other populations due to racial differences.

Conclusions

Overall, our study provides a hierarchy of evidence for the causal involvement of dysregulation of CXCL9, CCL11 and CASP8 in the pathogenesis of UC, from both observational and genetic analyses. Our study reveals that long-term pre-existing high levels of CXCL9, CXCL11 and CASP8 may increase the risk of UC without reverse causal interference. Our findings support the primary importance of dysregulation cytokines in the pathogenesis of UC and provide novel perspectives for treatment strategies and drug development based on mechanisms related to those cytokines and chemokines.

Acknowledgements

XL is supported by the Natural Science Fund for Distinguished Young Scholars of Zhejiang Province (LR22H260001) and the National Nature Science Foundation of China (grant no. 82204019). ET is supported by a CRUK Career Development Fellowship (C31250/A22804). KFD is supported by the project of the regional diagnosis and treatment center of the Health Planning Committee (No. JBZX-201903). SCL is supported by the Swedish Heart Lung Foundation (Hjärt-Lungfonden, 20210351), the Swedish Research Council (Vetenskapsrådet, 2019-00977), and the Swedish Cancer Society (Cancerfonden). This work is also funded by a grant to MGD as Project Leader with the MRC Human Genetics Unit Centre Grant (U127527198).

Conflict of interest

All authors declare that they have no conflict of interest.

Data availability statement

The analysis results of this study are included in this published article and its supplementary information files. The UK Biobank dataset can be accessed through their access application process.

References

1. Danese S, Fiocchi C. Ulcerative colitis. *N Engl J Med* 2011;**365**:1713-25.
2. Kaplan GG, Windsor JW. The four epidemiological stages in the global evolution of inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol* 2021;**18**:56-66.
3. Collaborators GBDIBD. 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. *The Lancet Gastroenterology & Hepatology* 2020;**5**:17-30.
4. Ng SC, Shi HY, Hamidi N*, et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: A systematic review of population-based studies. *Lancet (London, England)* 2017;**390**:2769-78.
5. Neurath MF. Cytokines in inflammatory bowel disease. *Nature Reviews Immunology* 2014;**14**:329-42.
6. Sanchez-Munoz F, Dominguez-Lopez A, Yamamoto-Furusho J-K. Role of cytokines in inflammatory bowel disease. *World Journal of Gastroenterology* 2008;**14**:4280-8.
7. Strober W, Fuss IJ. Proinflammatory cytokines in the pathogenesis of inflammatory bowel diseases. *Gastroenterology* 2011;**140**:1756-67.
8. Powrie F, Leach MW, Mauze S*, et al.* Inhibition of th1 responses prevents inflammatory bowel disease in scid mice reconstituted with cd45rbhi cd4+ t cells. *Immunity* 1994;**1**:553-62.
9. Neurath MF, Fuss I, Kelsall BL, Stüber E, Strober W. Antibodies to interleukin 12 abrogate established experimental colitis in mice. *The Journal of Experimental Medicine* 1995;**182**:1281-90.
10. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016;**40**:304-14.
11. Kalla R, Adams AT, Bergemalm D*, et al.* Serum proteomic profiling at diagnosis predicts clinical course, and need for intensification of treatment in inflammatory bowel disease. *J Crohns Colitis* 2021;**15**:699-708.
12. Bergemalm D, Andersson E, Hultdin J*, et al.* Systemic inflammation in preclinical ulcerative colitis. *Gastroenterology* 2021;**161**:1526-39.e9.
13. Zheng J, Haberland V, Baird D*, et al.* Phenome-wide mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* 2020;**52**:1122-31.
14. Suhre K, Arnold M, Bhagwat AM*, et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun* 2017;**8**:14357.
15. Sun BB, Maranville JC, Peters JE*, et al.* Genomic atlas of the human plasma proteome. *Nature* 2018;**558**:73-9.
16. Folkersen L, Gustafsson S, Wang Q*, et al.* Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab* 2020;**2**:1135-48.
17. Ferkingstad E, Sulem P, Atlason BA*, et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet* 2021;**53**:1712-21.
18. Liu JZ, van Sommeren S, Huang H*, et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics* 2015;**47**:979-86.
19. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through egger regression. *Int J Epidemiol* 2015;**44**:512-25.
20. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet* 2018;**50**:693-8.

21. Burgess S, Thompson SG. Bias in causal estimates from mendelian randomization studies with weak instruments. *Statistics in Medicine* 2011;**30**:1312-23.
22. Hemani G, Zheng J, Elsworth B*, et al.* The mr-base platform supports systematic causal inference across the human phenome. *Elife* 2018;**7**.
23. Elia G, Guglielmi G. Cxcl9 chemokine in ulcerative colitis. *La Clinica Terapeutica* 2018;**169**:e235-e41.
24. Liao F, Rabin RL, Yannelli JR*, et al.* Human mig chemokine: Biochemical and functional characterization. *The Journal of Experimental Medicine* 1995;**182**:1301-14.
25. Rotondi M, Chiovato L, Romagnani S, Serio M, Romagnani P. Role of chemokines in endocrine autoimmune diseases. *Endocrine Reviews* 2007;**28**:492-520.
26. Hakimizadeh E, Shamsizadeh A, Nazari M*, et al.* Increased circulating levels of cxc chemokines is correlated with duration and complications of the disease in type-1 diabetes: A study on iranian diabetic patients. *Clinical Laboratory* 2013;**59**:531-7.
27. Gonnella D. Type 1 diabetes and mig. *La Clinica Terapeutica* 2019;**170**:e465-e71.
28. James JA, Guthridge JM, Chen H*, et al.* Unique sjögren's syndrome patient subsets defined by molecular features. *Rheumatology (Oxford, England)* 2020;**59**:860-8.
29. Paparo SR. Rheumatoid arthritis and the th1 chemokine mig. *La Clinica Terapeutica* 2019;**170**:e472-e7.
30. Nakayama T, Yoshimura M, Higashioka K*, et al.* Type 1 helper t cells generate cxcl9/10-producing t-bet+ effector b cells potentially involved in the pathogenesis of rheumatoid arthritis. *Cellular Immunology* 2021;**360**:104263.
31. Midura-Kiela MT, Radhakrishnan VM, Larmonier CB*, et al.* Curcumin inhibits interferon-γ signaling in colonic epithelial cells. *American Journal of Physiology Gastrointestinal and Liver Physiology* 2012;**302**:G85-96.
32. Van Coillie E, Van Damme J, Opdenakker G. The mcp/eotaxin subfamily of cc chemokines. *Cytokine Growth Factor Rev* 1999;**10**:61-86.
33. Menzies-Gow A, Ying S, Sabroe I*, et al.* Eotaxin (ccl11) and eotaxin-2 (ccl24) induce recruitment of eosinophils, basophils, neutrophils, and macrophages as well as features of early- and late-phase allergic reactions following cutaneous injection in human atopic and nonatopic volunteers. *J Immunol* 2002;**169**:2712-8.
34. Mir A, Minguez M, Tatay J*, et al.* Elevated serum eotaxin levels in patients with inflammatory bowel disease. *Am J Gastroenterol* 2002;**97**:1452-7.
35. Coburn LA, Horst SN, Chaturvedi R*, et al.* High-throughput multi-analyte luminex profiling implicates eotaxin-1 in ulcerative colitis. *PLoS One* 2013;**8**:e82300.
36. Polosukhina D, Singh K, Asim M*, et al.* Ccl11 exacerbates colitis and inflammation-associated colon tumorigenesis. *Oncogene* 2021;**40**:6540-6.
37. Adar T, Shteingart S, Ben-Ya'acov A*, et al.* The importance of intestinal eotaxin-1 in inflammatory bowel disease: New insights and possible therapeutic implications. *Dig Dis Sci* 2016;**61**:1915-24.
38. Oberst A, Green DR. It cuts both ways: Reconciling the dual roles of caspase 8 in cell death and survival. *Nat Rev Mol Cell Biol* 2011;**12**:757-63.
39. Yeh WC, de la Pompa JL, McCurrach ME*, et al.* Fadd: Essential for embryo development and signaling from some, but not all, inducers of apoptosis. *Science* 1998;**279**:1954-8.
40. Moraes L, Magnusson MK, Mavroudis G*, et al.* Systemic inflammatory protein profiles distinguish irritable bowel syndrome (ibs) and ulcerative colitis, irrespective of inflammation or ibs-like symptoms. *Inflamm Bowel Dis* 2020;**26**:874-84.
41. Becker C, Watson AJ, Neurath MF. Complex roles of caspases in the pathogenesis of inflammatory bowel disease. *Gastroenterology* 2013;**144**:283-93.
42. El-Sherbiny M, Eisa NH, Abo El-Magd NF*, et al.* Anti-inflammatory/anti-apoptotic impact of betulin attenuates experimentally induced ulcerative colitis: An insight into tlr4/nf-kb/caspase signalling modulation. *Environ Toxicol Pharmacol* 2021;**88**:103750.

43. Satoh Y, Ishiguro Y, Sakuraba H*, et al.* Cyclosporine regulates intestinal epithelial apoptosis via tgf-beta-related signaling. *Am J Physiol Gastrointest Liver Physiol* 2009;**297**:G514-9.
44. Pierdomenico M, Negroni A, Stronati L*, et al.* Necroptosis is active in children with inflammatory bowel disease and contributes to heighten intestinal inflammation. *Am J Gastroenterol* 2014;**109**:279-87.
45. Lehle AS, Farin HF, Marquardt B*, et al.* Intestinal inflammation and dysregulated immunity in patients with inherited caspase-8 deficiency. *Gastroenterology* 2019;**156**:275-8.
46. Ananthakrishnan AN. Epidemiology and risk factors for ibd. *Nat Rev Gastroenterol Hepatol* 2015;**12**:205-17.
47. Ohlsson B. An okinawan-based nordic diet improves glucose and lipid metabolism in health and type 2 diabetes, in alignment with changes in the endocrine profile, whereas zonulin levels are elevated. *Exp Ther Med* 2019;**17**:2883-93.
48. Kohut ML, McCann DA, Russell DW*, et al.* Aerobic exercise, but not flexibility/resistance exercise, reduces serum il-18, crp, and il-6 independent of beta-blockers, bmi, and psychosocial factors in older adults. *Brain Behav Immun* 2006;**20**:201-9.

Table and Figure Legends

Table 1. Characteristics of two observatory studies of serum proteomic profiling data.

Abbreviations: UC, ulcerative colitis; IBD, inflammatory bowel disease.

Table 2. Results of the associations between ulcerative colitis and inflammatory proteins using reverse Mendelian Randomization analyses.

Abbreviations: SNPs, single nucleotide polymorphism; CI, confidence interval; SE, standard error; CXCL, C-X-C motif chemokine ligand; CCL, C-C motif chemokine ligand; CASP8, caspase 8; IVW, Inverse variance weighted.

Figure 1. Schematic diagram of the study design.

Abbreviations: UC, ulcerative colitis; IBD, inflammatory bowel disease; MR, Mendelian randomization; GWAS, genome-wide association study; FDR, false discovery rate; IVW, inverse variance weighted method; LD, linkage disequilibrium.

Figure 2. Adjusted results for analyses of serum proteomic data from discovery and validation datasets.

Abbreviations: UC, ulcerative colitis; MMP10, matrix metallopeptidase 10; CXCL, C-X-C motif chemokine ligand; CCL, C-C motif chemokine ligand; MCP1, i.e. CCL2; TGFA, transforming growth factor alpha; OSM, oncostatin M; HGF, hepatocyte growth factor; TNFSF, TNF superfamily member; IL, interleukin; ENRAGE, i.e. S100A12, S100 calcium binding protein A12; CASP8, caspase 8; Sig, significant.

Figure 3. Associations of 13 cytokines with UC risk in forward MR analysis.

Abbreviations: SNP, single nucleotide polymorphism; CI, confidence interval; CASP8, caspase 8; CCL, C-C motif chemokine ligand; MIP, major intrinsic protein of lens fiber; CXCL, C-X-C motif chemokine ligand; GRO, growth-regulated oncogene; IL, interleukin; MIG, monokine induced by

538 gamma interferon; ENRAGE, i.e. S100A12, S100 calcium binding protein A12; HGF, hepatocyte
539 growth factor; MMP10, matrix metallopeptidase 10; OSM, oncostatin M; TNFSF14, tumor necrosis
540 factor superfamily member 14; OR, odds ratio; CI, confidence interval.

Table 1. Characteristics of two observatory studies in secondary analysis of serum proteomic profiling data.

Clinical characteristics	Discovery cohort			Validation cohort			Pooled		
	UC (n=159)	Controls (n=293)	P value	UC (n=72)	Controls (n=140)	P value	UC (n=231)	Controls (n=433)	P value
Age, mean (SD)	37.3 (14.3)	32.4 (13.8)	<0.001	48.3 (10.8)	48.1 (10.7)	0.901	40.7 (14.2)	37.5 (14.9)	0.006
Sex, male, n (%)	92 (57.9)	133 (45.4)	0.015	34 (47.2)	64 (45.7)	0.950	126 (54.5)	197 (45.5)	0.032
Smoking status, n (%)									
Non-smoker	145 (91.2)	226 (77.1)	<0.001	50 (69.4)	105 (75.0)	0.241	195 (84.4)	331 (76.4)	0.003
Smoker	14 (8.8)	53 (18.1)		22 (30.6)	32 (22.9)		36 (15.6)	85 (19.6)	
Missing	0	14 (4.8)		0	3 (2.1)		0	17 (3.9)	
Disease extent for UC at diagnosis, n (%)									
Proctitis (E1)	39 (24.5)	-		16 (22.2)	-		55 (23.8)	-	
Left-sided colitis (E2)	47 (29.6)	-		28 (38.9)	-		75 (32.5)	-	
Extensive colitis (E3)	63 (39.6)	-		28 (38.9)	-		91 (39.4)	-	
Not available	10 (6.3)			0 (0)			10 (4.3)		

543 **Table 2. Associations between CXCL9, CCL11 and CASP8 levels and UC risk in the reverse MR analyses.**

Outcome	SNPs	Method	Beta coefficient (95% CI)	SE	P $_{Effect}$	P $_{Heterogenicity}$	P $_{Intercept}$
CXCL9	42	IVW	0.01 (-0.01, 0.02)	0.007	0.296	0.390	-
		MR Egger	0.02 (-0.02, 0.05)	0.016	0.347	0.361	0.587
		Weighted median	0.01 (-0.01, 0.03)	0.010	0.250	-	-
		MR-PRESSO	0.01 (-0.01, 0.02)	0.007	0.405	-	-
CCL11	42	IVW	0.01 (0, 0.03)	0.009	0.115	0.002	-
		MR Egger	0.02 (-0.02, 0.06)	0.021	0.398	0.001	0.860
		Weighted median	0.02 (-0.01, 0.04)	0.011	0.142	-	-
		MR-PRESSO	0.02 (0, 0.03)	0.008	0.002	-	-
CASP8	42	IVW	0.001 (-0.01, 0.01)	0.007	0.978	0.380	-
		MR Egger	0.01 (-0.02, 0.04)	0.016	0.529	0.359	0.490
		Weighted median	0.01 (-0.01, 0.03)	0.010	0.394	-	-
		MR-PRESSO	0 (-0.01, 0.01)	0.007	0.325	-	-

544

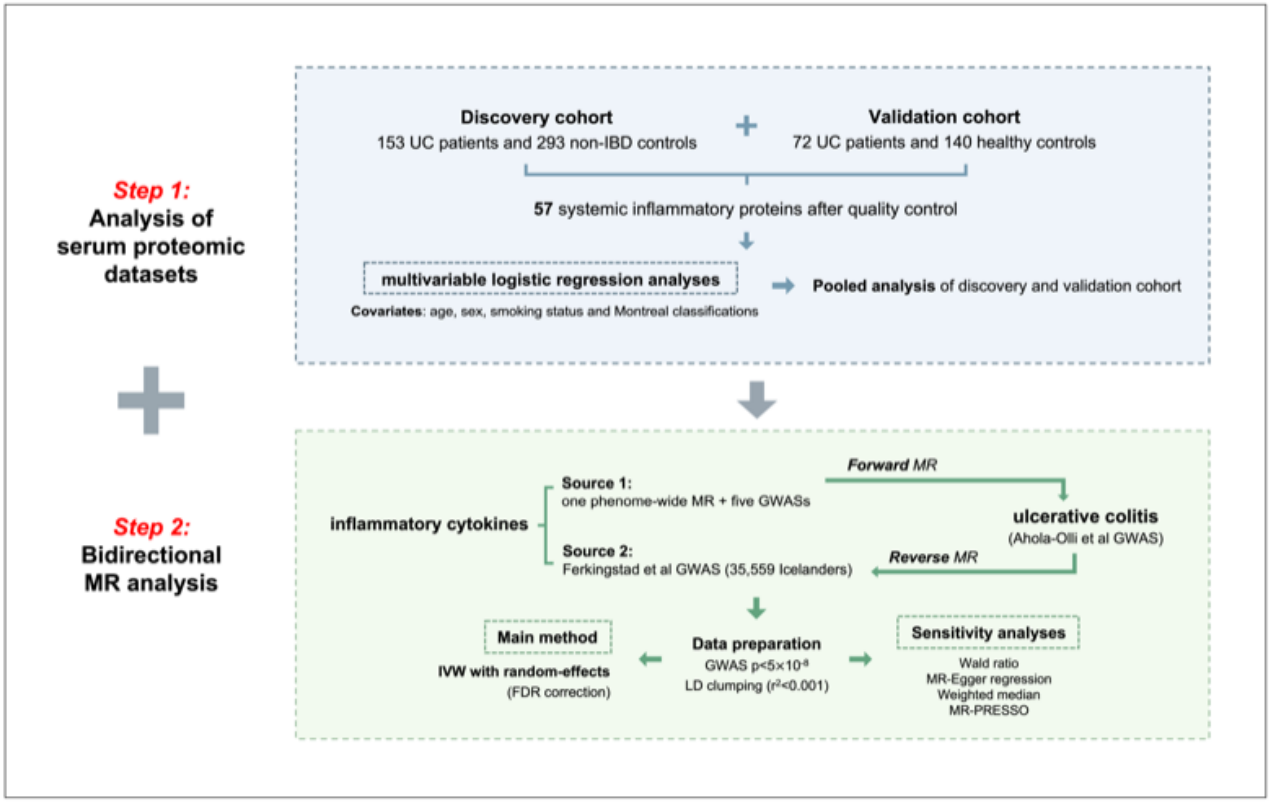


Figure 1. Schematic diagram of the study design.

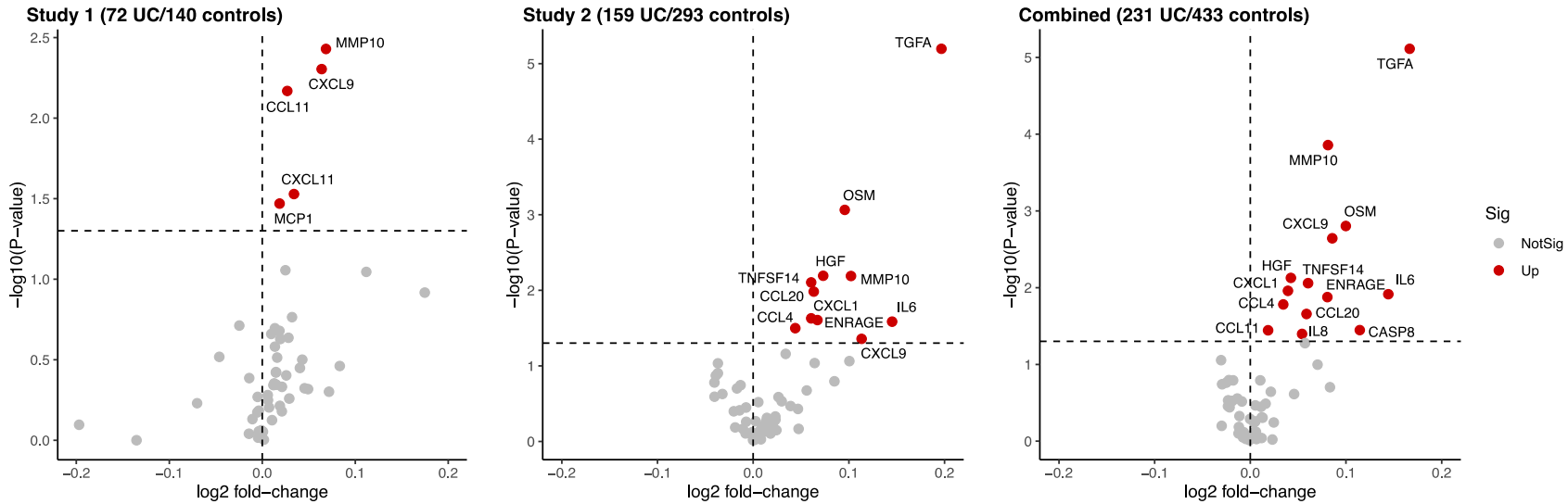


Figure 2. Adjusted results for analyses of serum proteomic data from discovery and validation datasets.

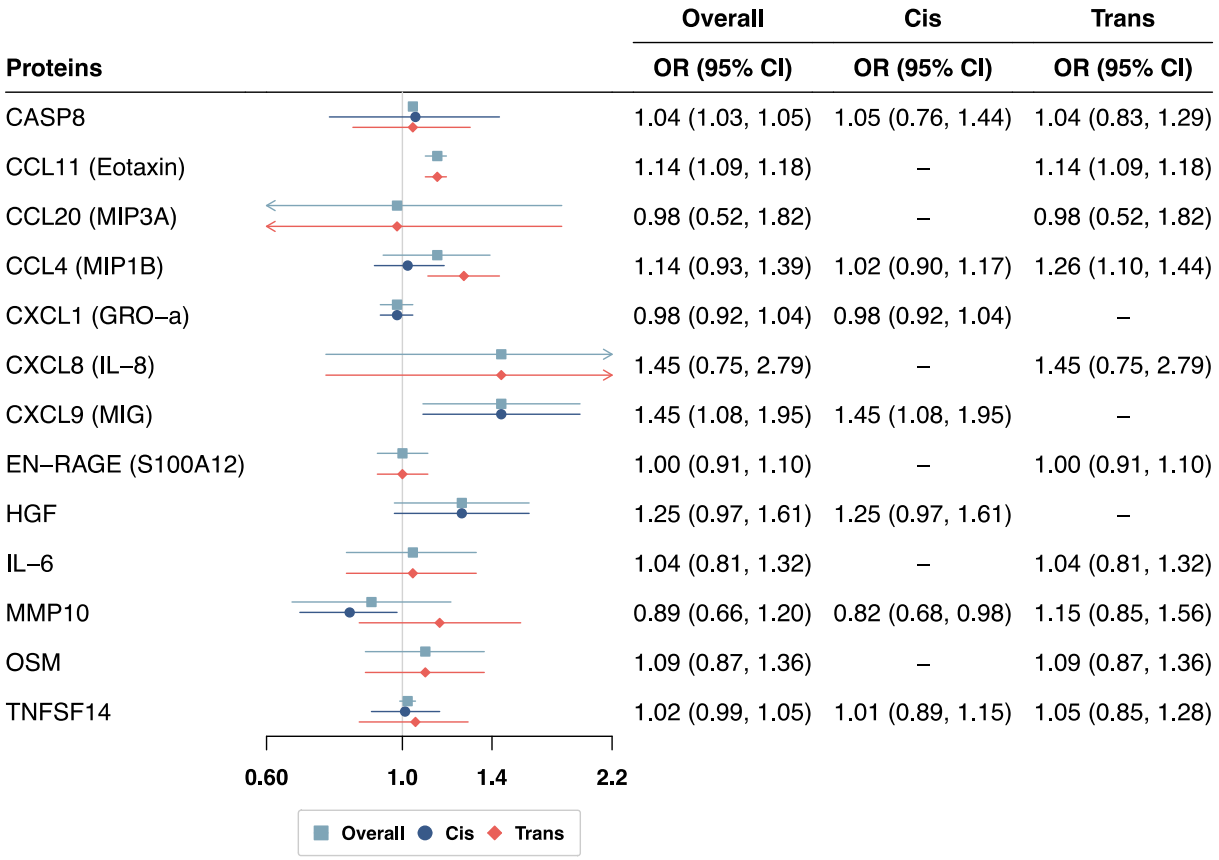


Figure 3. Associations of 13 cytokines with UC risk in forward MR analysis.