RESEARCH ARTICLE

Oxidative stress gene expression, DNA methylation, and gut microbiota interaction trigger Crohn's disease: a multi-omics Mendelian randomization study

Shu Xu[1†], Xiaozhi Li[1†], Shenghong Zhang[1], Cancan Qi[2], Zhenhua Zhang[3], Ruiqi Ma[1], Liyuan Xiang[1], Lianmin Chen[4,5], Yijun Zhu[1,6], Ce Tang[1,6], Arno R. Bourgonje[7], Miaoxin Li[8], Yao He[1], Zhirong Zeng[1], Shixian Hu[1,6*†], Rui Feng[1,9*†] and Minhu Chen[1*†]

Abstract

Background Oxidative stress (OS) is a key pathophysiological mechanism in Crohn's disease (CD). OS-related genes can be affected by environmental factors, intestinal inflammation, gut microbiota, and epigenetic changes. However, the role of OS as a potential CD etiological factor or triggering factor is unknown, as differentially expressed OS genes in CD can be either a cause or a subsequent change of intestinal inflammation. Herein, we used a multi-omics summary data-based Mendelian randomization (SMR) approach to identify putative causal effects and underlying mechanisms of OS genes in CD.

Methods OS-related genes were extracted from the GeneCards database. Intestinal transcriptome datasets were collected from the Gene Expression Omnibus (GEO) database and meta-analyzed to identify differentially expressed genes (DEGs) related to OS in CD. Integration analyses of the largest CD genome-wide association study (GWAS) summaries with expression quantitative trait loci (eQTLs) and DNA methylation QTLs (mQTLs) from the blood were performed using SMR methods to prioritize putative blood OS genes and their regulatory elements associated with CD risk. Up-to-date intestinal eQTLs and fecal microbial QTLs (mbQTLs) were integrated to uncover potential interactions between host OS gene expression and gut microbiota through SMR and colocalization analysis. Two additional Mendelian randomization (MR) methods were used as sensitivity analyses. Putative results were validated in an independent multi-omics cohort from the First Affiliated Hospital of Sun Yat-sen University (FAH-SYS).

Results A meta-analysis from six datasets identified 438 OS-related DEGs enriched in intestinal enterocytes in CD from 817 OS-related genes. Five genes from blood tissue were prioritized as candidate CD-causal genes using

*Correspondence:
Shixian Hu
hushx9@mail.sysu.edu.cn
Rui Feng
fengr7@mail.sysu.edu.cn
Minhu Chen
chenminhu@mail.sysu.edu.cn
Full list of author information is available at the end of the article

Xu *et al. BMC Medicine*   (2023) 21:179

Page 2 of 16

three-step SMR methods: *BAD*, *SHC1*, *STAT3*, *MUC1*, and *GPX3*. Furthermore, SMR analysis also identified five putative intestinal genes, three of which were involved in gene–microbiota interactions through colocalization analysis: *MUC1*, *CD40*, and *PRKAB1*. Validation results showed that 88.79% of DEGs were replicated in the FAH-SYS cohort. Associations between pairs of *MUC1*–*Bacillus aciditolerans* and *PRKAB1*–*Escherichia coli* in the FAH-SYS cohort were consistent with eQTL–mbQTL colocalization.

Conclusions This multi-omics integration study highlighted that OS genes causal to CD are regulated by DNA methylation and host-microbiota interactions. This provides evidence for future targeted functional research aimed at developing suitable therapeutic interventions and disease prevention.

Keywords Oxidative stress, Crohn's disease, Integrative omics, Mendelian randomization

Background

Crohn's disease (CD) is a type of chronic and relapsing inflammatory bowel disease (IBD) that affects the gastrointestinal tract and is accompanied by extraintestinal manifestations and perianal diseases [1]. Although the etiology of CD remains unclear, a complex interplay between genetic variation, environmental factors, immune dysfunction, and intestinal microbiota is believed to underlie disease pathogenesis [2]. Unraveling the complexity behind this interplay may provide crucial insights into CD pathogenesis and expose potential targets for therapeutic interventions and disease prevention.

Oxidative stress (OS) is defined as an imbalance between oxidants and antioxidants in favor of the oxidants leading to a disruption of redox signaling and control and/or molecular damage [3]. Multiple OS-related genes contribute to the complex multifactorial pathophysiology in CD [4, 5]. For example, genetic polymorphisms of the inducible nitric oxide synthase gene (encoded by *NOS2A*) are associated with IBD susceptibility accompanied by increased gene expression, suggesting a vital role of genetic effects on OS genes in CD [6]. The nicotinamide adenine dinucleotide phosphate oxidase genes *NOX1* and *DUOX2* also play a key role in mediating reactive oxygen species (ROS) generation. Overexpression of these genes is involved in impaired intestinal barrier integrity, microbial dysbiosis, and bacterial invasion, highlighting the association between host OS signaling and gut microbiota in CD [7–10]. Moreover, DNA methylation (DNAm) modulates redox homeostasis by regulating gene expressions of *NRF2*, *HIF1A*, and related proteins [11–14]. However, few studies have addressed whether OS has a causative role in triggering CD or merely inflicts collateral tissue damage alongside intestinal inflammation. Studying the underlying disease mechanisms of OS-related genes may help identify potential pathogenetic factors and redox-related therapeutic targets for IBD [15].

Although a growing number of studies have suggested relevant OS genes in CD, no study has comprehensively and systematically identified their potential causal association with this disease. Genome-wide association studies (GWASs) have been employed to identify genomic loci containing OS genes associated with CD [16, 17]. However, the top associated variants may not be causal because of the complicated linkage disequilibrium (LD) structure of genomes [18, 19]. Moreover, these genetic variants can potentially regulate DNAm, gene expression, protein levels, and the abundance of gut microbiota [20, 21]. Integration of multi-omics is an emerging approach in the post-GWAS era to identify critical regulators for exploring therapeutic targets in CD [22]. For example, summary data-based Mendelian randomization (SMR) that integrates IBD GWAS data with expression quantitative trait loci (eQTLs) has been developed to prioritize causal variants mediated by gene expression in the blood [20]. However, the causal OS genes in CD-affected tissues and their interactions with gut microbiota are poorly understood [23–25].

This study presents a multi-omics-based Mendelian randomization (MR) study to identify the putative causal effects and molecular mechanisms of OS genes in CD using blood and intestinal tissues. A sizable intestinal transcriptome meta-analysis was performed to identify differentially expressed CD-related OS genes. Utilizing SMR methods, we integrated the largest CD GWAS summary statistics with eQTLs and DNA methylation QTLs (mQTLs) in the blood. Furthermore, up-to-date intestinal eQTLs and fecal microbial QTLs (mbQTLs) were first integrated into the current analysis to uncover the potential interactions between host OS genes and gut microbiota. Two additional MR methods were used as sensitivity analyses to test the heterogeneity. Finally, the putative results were then partially replicated in an independent multi-omics cohort.

Methods

Study design and data resources

Figure 1 describes the design of this study. OS-related genes were extracted from the GeneCards database

Xu *et al. BMC Medicine* (2023) 21:179

Page 3 of 16

(v5.10, https://www.genecards.org) using the keyword "oxidative stress" with a relevance score ≥ 7 according to previous methods [26–28]. Six publicly available transcriptome datasets containing intestinal biopsies from patients with CD and healthy controls (HCs) were obtained from the Gene Expression Omnibus (GEO) database [29–39] and meta-analyzed to identify differentially expressed genes (DEGs) related to OS in CD. GWAS summary statistics for CD were derived from a meta-analysis of two separate IBD GWASs, yielding a sample size of 12,194 patients with CD and 28,072 HCs based on European populations [40]. Blood eQTL summary statistics of OS genes were obtained from eQTLGen including the genetic data of blood gene expression in 31,684 individuals derived from 37 datasets [41]. Blood mQTL summary data were generated from a meta-analysis of two European cohorts: the Brisbane Systems Genetics Study ($n = 614$) and the Lothian Birth Cohorts ($n = 1366$) [19]. Intestinal eQTL data were from the Genotype-Tissue Expression (GTEx) project ($n = 860$) [42] and the 1000IBD cohorts ($n = 299$) [43]. The current study focused only on *cis*-eQTLs and *cis*-mQTLs, which constituted single nucleotide polymorphisms (SNPs) within a 1-Mb distance from the start and end of the gene. Fecal mbQTL data was generated from the Dutch Microbiome Project (DMP) study, which included data from 7738 individuals to assess the host genetic effects on the gut microbiota [44].

For external validation, 46 treatment-naïve patients with CD and 44 HC subjects were prospectively recruited from the First Affiliated Hospital of Sun Yat-sen University (FAH-SYS) IBD multi-omics cohort. Paired intestinal biopsies and stool specimens were collected for RNA sequencing (RNA-seq) and shotgun metagenomic sequencing, respectively.

More detailed information on study datasets for this study is provided in Additional file 1: Supplementary Methods [19, 29–45].

Statistical analysis

Meta-analysis of DEGs

Intestinal DEGs related to OS in patients with CD and HCs were analyzed using a linear regression model while adjusting for age, sex, body mass index, and medication usage if metadata were available. To increase statistical power, we pooled ileal and colonic biopsies and added tissue location as a covariate in linear models. DEGs were analyzed separately in six gene expression datasets, followed by a fixed-effects meta-analysis using the R package *metafor*.

Meta-analysis of intestinal cis-eQTLs

To include as many intestinal *cis*-eQTLs as possible, we first performed a meta-analysis using the meta-analysis of *cis*-eQTL in the correlated sample (MeCS) method across the transverse colon, sigmoid colon, and small intestine of the GTEx results for OS genes, considering the sample overlaps. The intestinal *cis*-eQTLs from the 1000IBD cohorts were highly comparable to those detected in the non-disease GTEx dataset (consistency rates > 97%) [43]. Therefore, we used the conventional inverse-variance-weighted meta-analysis in SMR (v1.3.1) for the two independent datasets.

SMR and colocalization analysis

SMR multi-tools have been established to detect whether the effect of SNPs on the phenotype is mediated by molecular traits such as gene expression, DNAm, and gut microbiota. Colocalization analysis aimed to investigate the overlapping variants likely responsible for different traits. The integration of data from GWAS with other molecular QTL data by SMR or colocalization improved the detection of candidate causal SNPs via specific pathways.

Blood tissue analysis used the SMR multi-tool to determine the causal inference of OS genes and the 1000 Genomes European reference to calculate LD. A three-step SMR analysis was performed: (1) SNPs were

(See figure on next page.)

Fig. 1 Workflow of the study. A series of analyses was conducted to identify candidate causal oxidative stress (OS) genes associated with Crohn's disease (CD) onset. OS-related genes were extracted from the GeneCards database. Six intestinal transcriptome datasets including patients with CD and healthy controls (HCs) were obtained from the GEO database and meta-analyzed to identify differentially expressed CD-related OS genes, followed by cell type-specific expression analysis (CSEA). Integration of GWAS summaries and *cis*-eQTLs/*cis*-mQTLs data from the blood by using three-step SMR methods prioritized putative blood OS genes and their regulatory elements associated with the risk of CD (SMR FDR < 0.05; HEIDI test $P > 0.05$). Sensitivity analyses were performed after the primary SMR to test the heterogeneity (Cochran Q statistic implemented in MR-Egger and inverse variance weighting (IVW) method, $P > 0.05$ indicates no heterogeneity exists). Moreover, we meta-analyzed the intestinal *cis*-eQTLs from two public summaries (GTEx and 1000IBD) and further integrated the meta-intestinal *cis*-eQTLs with fecal mbQTLs from the Dutch Microbiome Project (DMP) to uncover the potential interactions between OS genes and gut microbiota through SMR, sensitivity, and colocalization analysis (SMR FDR < 0.05; HEIDI test $P > 0.05$; Cochran Q $P > 0.05$; colocalization PPH4 > 0.5). An external First Affiliated Hospital of Sun Yat-sen University (FAH-SYS) inflammatory bowel disease (IBD) multi-omics cohort with paired intestinal bulk RNA-seq and fecal metagenomics data was used to validate the differentially expressed genes (DEGs). The directional associations between colocalized gene-microbiota were investigated as complementary evidence of OS gene-microbiota interactions
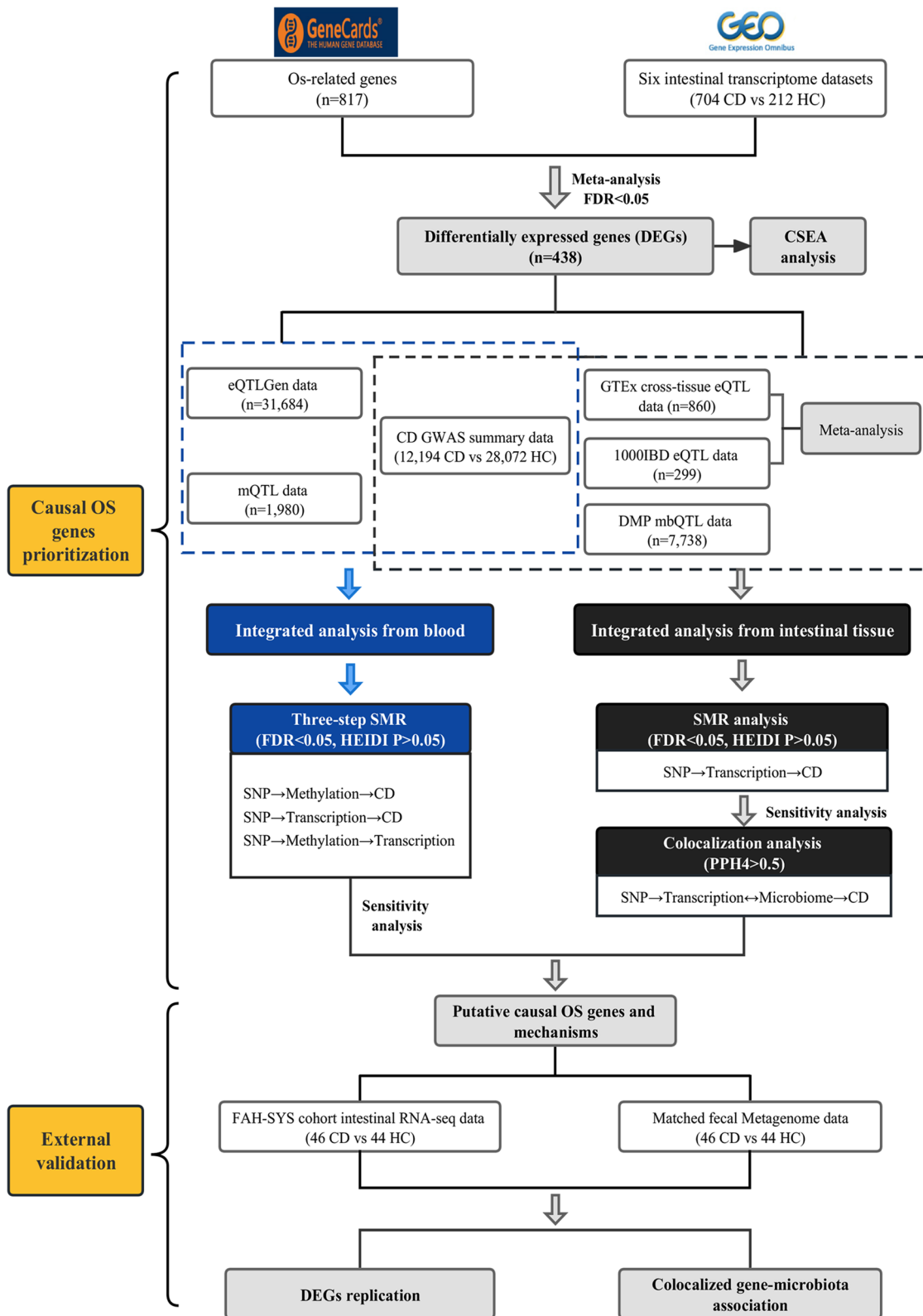
Xu *et al. BMC Medicine* (2023) 21:179

Page 4 of 16

Fig. 1 (See legend on previous page.)

Xu *et al. BMC Medicine*   (2023) 21:179

Page 5 of 16

instruments, blood gene expressions were exposure, and CD was the outcome; (2) SNPs were instruments, blood DNAms were exposure, and CD was the outcome; (3) SNPs were instruments, blood DNAms were exposure, and blood gene expressions were the outcome. The third step included only significant signals from steps 1 and 2. The final candidate signals were defined as those that (1) passed all three-step SMR false discovery rate (FDR) < 0.05; (2) were suggestively significant genome-wide ($P < 1 \times 10^{-5}$) in all eQTLs, mQTLs, and GWAS; and (3) exhibited heterogeneity in the dependent instrument (HEIDI) test results with $P > 0.05$.

Intestinal tissue analysis used the SMR multi-tool to determine the causal inference between GWAS and *cis*-eQTLs. This included SNPs as instruments, intestinal gene expressions as exposure, and CD as the outcome (SMR FDR < 0.05, HEIDI $P > 0.05$, *cis*-eQTLs, and GWAS $P < 1 \times 10^{-5}$).

Sensitivity analyses were conducted after completing the primary SMR analyses with two additional MR methods. We tested for heterogeneity across the individual causal effects using Cochran Q statistic implemented in both MR–Egger and inverse variance weighting (IVW) methods. The P value of Cochran's Q test < 0.05 or the HEIDI < 0.05 indicates the existence of heterogeneity.

Colocalization was chosen to assess the potential interactions between intestinal gene expression and microbiota because of the limited power to assess the causality between gut microbiota and diseases [46, 47]. It is a method to assess the presence of a shared causal variant in the region for two traits. Analysis was performed using the *coloc* R package with PPH4 > 0.5 as the threshold for the shared genetic effects between the two traits [48, 49].

Replication in the FAH-SYS IBD multi-omics cohort
(1) Significant DEG results from the six-dataset meta-analysis were selected to test between patients with CD and healthy controls in our cohort, and (2) the significant gene expression–microbial pathway pairs from the colocalization analysis were selected. The correlations between intestinal gene expression and pathway-related microbial taxa (genus- and species-level) abundance were then assessed using a linear regression model adjusted for age, sex, and body mass index. Pathway-related microbial taxa were defined using MetaCyc (v24.0) *pathways_to_organisms* definitions (https://metacyc.org/) containing the potential microbiota involved in each pathway. Only taxa with a present rate > 10% in the FAH-SYS IBD cohort were kept for the analysis resulting in three species and two genera. More specifically, CRNFORCAT.PWY was mainly predicted from *Methylobacterium* and *Cupriavidus*, P164.PWY

and PYRIDNUCSYN.PWY were mainly predicted from *Escherichia coli*, PWY.5101 was mainly predicted from *E. coli* and *Saccharomyces cerevisiae*, and PWY.7237 was mainly predicted from *E. coli* and *Bacillus aciditolerans*, while no PWY0.845-associated taxa survived after filtering. A P value < 0.05 was considered as the significant threshold considering the small sample size of the FAH-SYS IBD cohort.

Unless otherwise stated, statistical significance was defined as FDR < 0.05 by the Benjamini–Hochberg (BH) method.

Cell type-specific enrichment and regulatory component annotation
The Cell type-Specific Enrichment Analysis DataBase (CSEA-DB, https://bioinfo.uth.edu/CSEADB/) was used to investigate whether intestinal DEGs were specific to any cell type. Among 126 general cell types from 111 tissues, we focused on cell types present in the intestine, which resulted in seven tissues and 23 general cell types, after multiple testing corrections (BH method, FDR < 0.05).

The regulatory signature enrichment of DNAm sites was assessed using eFORGE (http://eforge.cs.ucl.ac.uk/), including chromatin status (active and inactive) and histone marker (H3K4me1 and H3K4me3) annotation. The regions of the individual DNAm sites were annotated at http://grch37.ensembl.org/.

Results
Meta-analysis of differentially expressed OS genes between patients with CD and HCs
To understand the role of OS genes in CD, a total of six gene expression datasets (three microarray datasets and three RNA-seq datasets) were included to compare RNA expression in the intestinal tissues of patients with CD ($n = 704$) and HCs ($n = 212$) through a meta-analysis (Additional file 2: Table S1). In total, 817 OS-related genes with a relevance score ≥ 7 were downloaded from GeneCards (the "Methods" section; Additional file 2: Table S2). Subsequently, 438 OS genes were differentially expressed between CD and control tissues (FDR < 0.05) according to the meta-analysis (Fig. 2A and Additional file 2: Table S3). The top five genes prioritized by effect size were *DUOX2*, *MMP3*, *S100A8*, *MMP1*, and *IL1B*, which are all reported to be associated with intestinal inflammation in IBD [7, 50–52]. Furthermore, we conducted cell type-specific expression analysis (CSEA) of these DEGs. OS-related DEGs were significantly enriched in intestinal enterocytes (FDR = 0.002) among the 23 general cell classifications (Fig. 2B; the "Methods" section; Additional file 2: Table S4), suggesting an

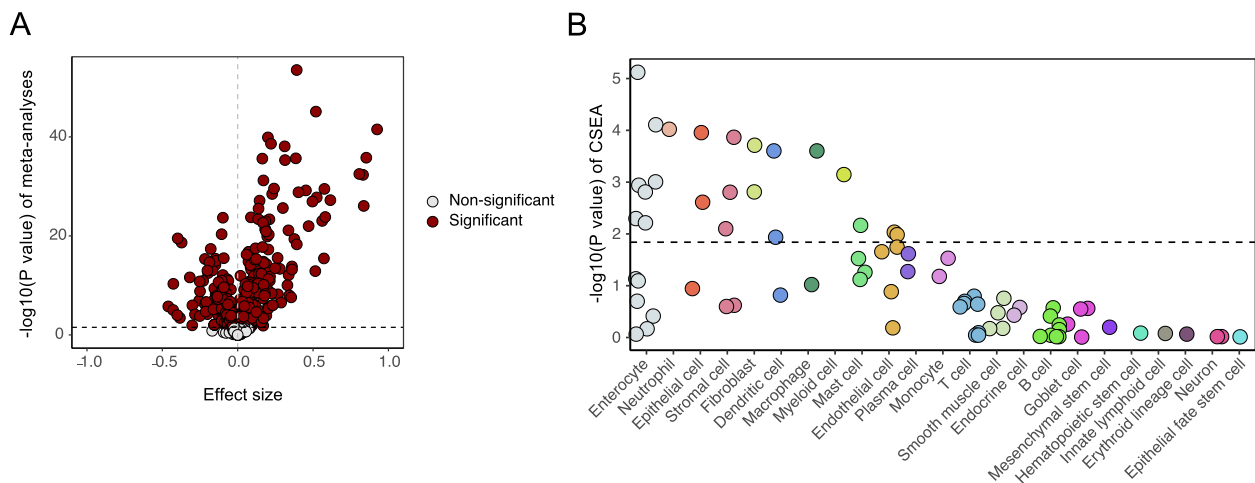Xu *et al. BMC Medicine*   (2023) 21:179

Page 6 of 16

Fig. 2 Meta-analysis of six intestinal gene expression datasets between patients with CD and HCs. **A** In total, 708 out of 817 genes presented in all six intestinal transcriptome data were assessed for expression differences between patients with CD and HCs. The volcano plot shows the meta effect sizes on the *x*-axis while the *y*-axis indicates the −log10-transformed meta *P* values. Red dots are the 438 significant differentially expressed genes (DEGs), and gray dots represent non-significantly expressed genes. The dashed line indicates the significant threshold with FDR < 0.05 corrected for the number of gene tests. **B** Cell type-Specific Enrichment Analysis DataBase was used to investigate whether the intestinal DEGs were specific to any cell type in the small intestine and colon. The *x*-axis indicates the cell types derived from the intestinal tissue and blood. Dots represent 77 small intestine and colon cell types annotated by 23 general classifications descending by order of significance. The dashed line is the significant threshold with FDR < 0.05

essential role of the epithelial cells in regulating intestinal OS and maintaining mucosal homeostasis.

Integration of GWAS and OS-related eQTL/mQTL data from the blood

Given such an amount of OS DEGs (53.61%, 438 out of 817) observed in patients with CD compared with HCs, we then hypothesized that gene expression may explain the plausible causality in the disease. Additionally, DNAm located in promoters or enhancers commonly influences the regulation of disease-associated target genes [19]. Therefore, we aimed to identify candidate causal genes for CD and explore their possible underlying epigenetic mechanism of gene regulation in the blood. A three-step SMR method was used, and only the significant results in all three SMR analyses that passed sensitivity checks were interpreted as suggestive causal genes (the "Methods" section). Herein, 438 OS-related DEG *cis*-eQTLs (94,037 SNP–gene pairs) and their *cis*-mQTLs (52,761 SNP–CpG sites) were integrated with the largest available GWAS summary statistics for CD.

In concrete, the integration of eQTL results from the eQTLGen Consortium (*n* = 31,684) and CD GWAS summary statistics resulted in 16 OS-related genes (FDR < 0.05, HEIDI *P* > 0.05, and Cochran's *Q P* > 0.05) (Additional file 2: Table S5). Meanwhile, we identified 665 DNAm probes (near 127 genes within 1 Mb) by integrating the same CD GWAS results and mQTL summary statistics from the meta-analysis of Brisbane

Systems Genetics Study and Lothian Birth Cohorts (*n* = 1980) (Additional file 2: Table S6). Further integration analysis of putative CD-causal *cis*-eQTL and *cis*-mQTL data prioritized eight DNAm probes potentially regulating five neighboring genes: *BAD*, *SHC1*, *STAT3*, *MUC1*, and *GPX3* (SMR FDR < 0.05, HEIDI *P* > 0.05, and Cochran's *Q P* > 0.05) (Additional file 2: Table S7). As expected, these CpG sites were significantly enriched in the transcription start sites (TSSs) of peripheral blood cells, including primary hematopoietic stem cells (FDR = 1.65×10^{-16}), primary T helper memory blood cells (FDR = 1.94×10^{-15}), and primary mononuclear blood cells (FDR = 4.79×10^{-12}) (top three significance) (Additional file 3: Fig. S1; Additional file 2: Table S8).

Putative CD-causal genes mediated by blood methylation regulation on gene expression

Our three-step SMR analysis prioritized *STAT3*, a well-known redox-regulated gene, and its expression is strictly affected by the intracellular redox environment [53]. This study showed that the SNP signals associated with *STAT3* were significant across the data from CD GWAS, eQTL, and mQTL studies. The DNAm probe cg06422947 was found to be located in the enhancer region, 427 kbp upstream of *STAT3*. The methylation level of this site showed a negative effect on *STAT3* expression (beta_{SMR} = − 0.11) and CD onset (beta_{SMR} = − 0.09), while the *STAT3* expression level was positively associated with

Xu *et al. BMC Medicine*   (2023) 21:179

Page 7 of 16

the disease (beta$_{SMR}$ = 0.70). Together, our results suggest a putative mechanism wherein a lower DNAm level at the enhancer region of *STAT3* upregulates the expression of *STAT3* and subsequently increases CD risk (Fig. 3A, B).

Another key example is *GPX3* (Fig. 3C, D) which belongs to the glutathione peroxidase family and catalyzes the reduction of organic hydroperoxides and hydrogen peroxide by glutathione, thereby protecting cells against oxidative damage [54]. We found that DNAm probe cg08580836, located in the promoter region, was causally negatively associated with *GPX3* expression (beta$_{SMR}$ = −0.24). Consistently, higher *GPX3* gene expression (beta$_{SMR}$ = −0.15) and lower methylation levels (beta$_{SMR}$ = 0.74) potentially decreased the risk of CD. Thus, the putative mechanism could be that the genetic variants upregulate *GPX3* expression by influencing the promoter DNAm status, showing a protective effect on CD onset.

Integration of GWAS and OS-eQTL/mbQTL data from intestinal tissue

Genetic effects on gene expression vary across blood and intestine tissues, which could reflect different CD-causal genes [55]. Moreover, host genetics and gut microbiota are known to play critical roles in CD [56]. Intestinal tissues are in direct contact with gut microbes and sense local changes in OS levels; therefore, we hypothesized that integrating *cis*-eQTLs and mbQTLs from intestinal tissue would provide novel candidate targets with putative host–microbiota interactions. First, we performed a meta-analysis of *cis*-eQTL data from three intestinal tissues (sigmoid colon, transverse colon, and small intestine) obtained from the GTEx project (*n* = 860) adjusting for sample overlap. This was followed by a meta-analysis with intestinal *cis*-eQTL data from the 1000IBD project (*n* = 299). The final meta-analysis identified 43,975 *cis* SNP–gene pairs corresponding to 392 OS-related genes (FDR < 0.05) (Additional file 2: Table S9). The SMR analysis demonstrated the potential causal role of five intestinal-expressed genes in CD (SMR FDR < 0.05, HEIDI *P* > 0.05, and Cochran's *Q P* > 0.05): *MUC1, CD40, PARK7, PRKAB1*, and *NDUFS1* (Additional file 2: Table S10).

To further explore the role of intestinal OS genes from the perspective of host–microbiota interactions, we integrated mbQTL summary statistics with putative CD-causal *cis*-eQTLs by colocalization analysis.

This analysis was assumed to determine the probability that the genetic determinants of mucosal gene expression were shared with gut microbiota. SMR was not used because of power issues suffering from the moderate effects of host genetics on the gut microbiota [46, 47]. In total, six gene expression–microbial pathway pairs were detected at the threshold of PPH4 > 0.5, including three of the above genes, *MUC1, CD40*, and *PRKAB1* (Additional file 2: Table S11; Additional file 4: Fig. S2).

Putative CD-causal genes involved in intestinal gene–microbiota interactions

We prioritized *MUC1* as a candidate OS causal gene in CD intestinal tissues associated with the gut microbiota based on SMR and colocalization analysis (Fig. 4A). Our study showed that elevated *MUC1* expression likely played a causal role in CD onset (beta$_{SMR}$ = 0.57). Furthermore, SNPs regulating *MUC1* expression might also affect the microbial metabolic functions given the colocalization analysis. Concretely, the creatinine degradation (CRNFORCAT.PWY; PPH4 = 0.94) and myo-inositol degradation (PWY.7237; PPH4 = 0.60) microbiota metabolic pathways, which are associated with decreased inflammation or short-chain fatty acids production, shared genetic effects with *MUC1* expression. Our findings suggest that genetic variation in *MUC1* might simultaneously regulate its gene expression and the production of microbiota-derived metabolites, thus increasing the risk of CD.

The *CD40* gene for CD is another example involved in intestinal gene–microbiota interactions (Fig. 4B). *CD40* and its ligand (*CD40L*) are associated with ROS production in immune and endothelial cells [57, 58]. They constitute the second activation signal of macrophages and enhance its activation and the resultant production of potential antimicrobial peptides as well as ROS, nitric oxide, and related metabolites [59]. This study showed a protective effect of CD40 expression against CD (beta$_{SMR}$ = −0.62), which was also found in the case of other immune-related disease studies [60]. Three microbial pathways were associated with *CD40* gene expression under genetic regulation: nicotinamide adenine dinucleotide biosynthesis from aspartate (PYRIDNUCSYN.PWY; PPH4 = 0.83), L-isoleucine biosynthesis (PWY.5101; PPH4 = 0.72), and the superpathway of pyridoxal 5′-phosphate biosynthesis and salvage

(See figure on next page.)

Fig. 3 Three-step SMR analysis prioritized putative causal OS genes and mechanisms in CD using blood tissue. Examples of well-known and novel CD-causal OS genes. **A, C** Locus zoom plots showing the consistent genetic effects from CD GWAS, *cis*-mQTL, and *cis*-eQTL nearby *STAT3* and *GPX3* (from upper to lower panels, all minimum $P < 1 \times 10^{-5}$). **B, D** Three-step SMR indicating significant causal relationships between gene expressions and CD onset mediated by methylation (all three-step SMR FDR < 0.05, HEIDI test *P* > 0.05). From left to right: SMR between gene expression and CD GWAS, SMR between gene methylation and CD GWAS, and SMR between gene methylation and expression
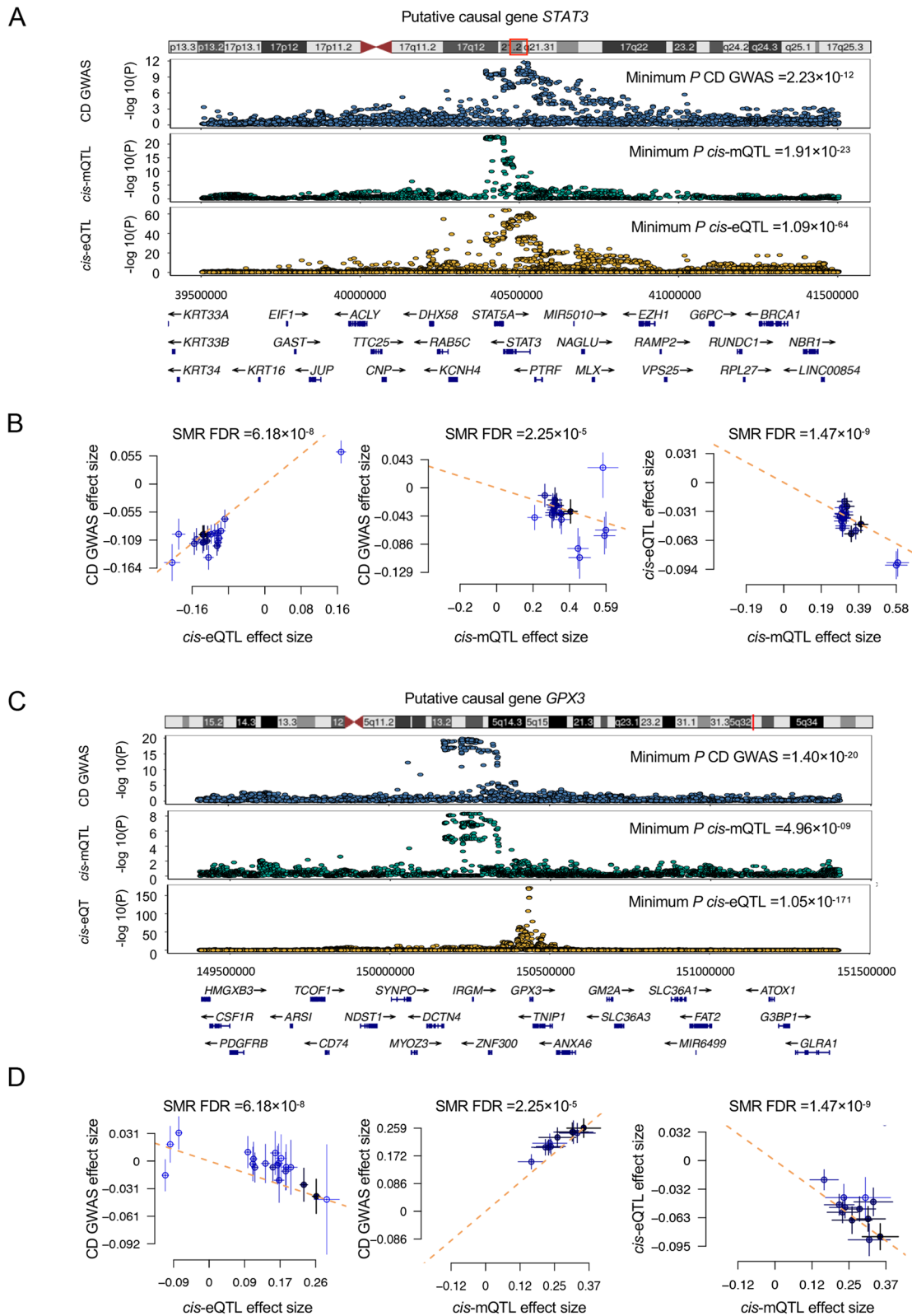
Xu *et al. BMC Medicine* (2023) 21:179

Page 8 of 16

Fig. 3 (See legend on previous page.)

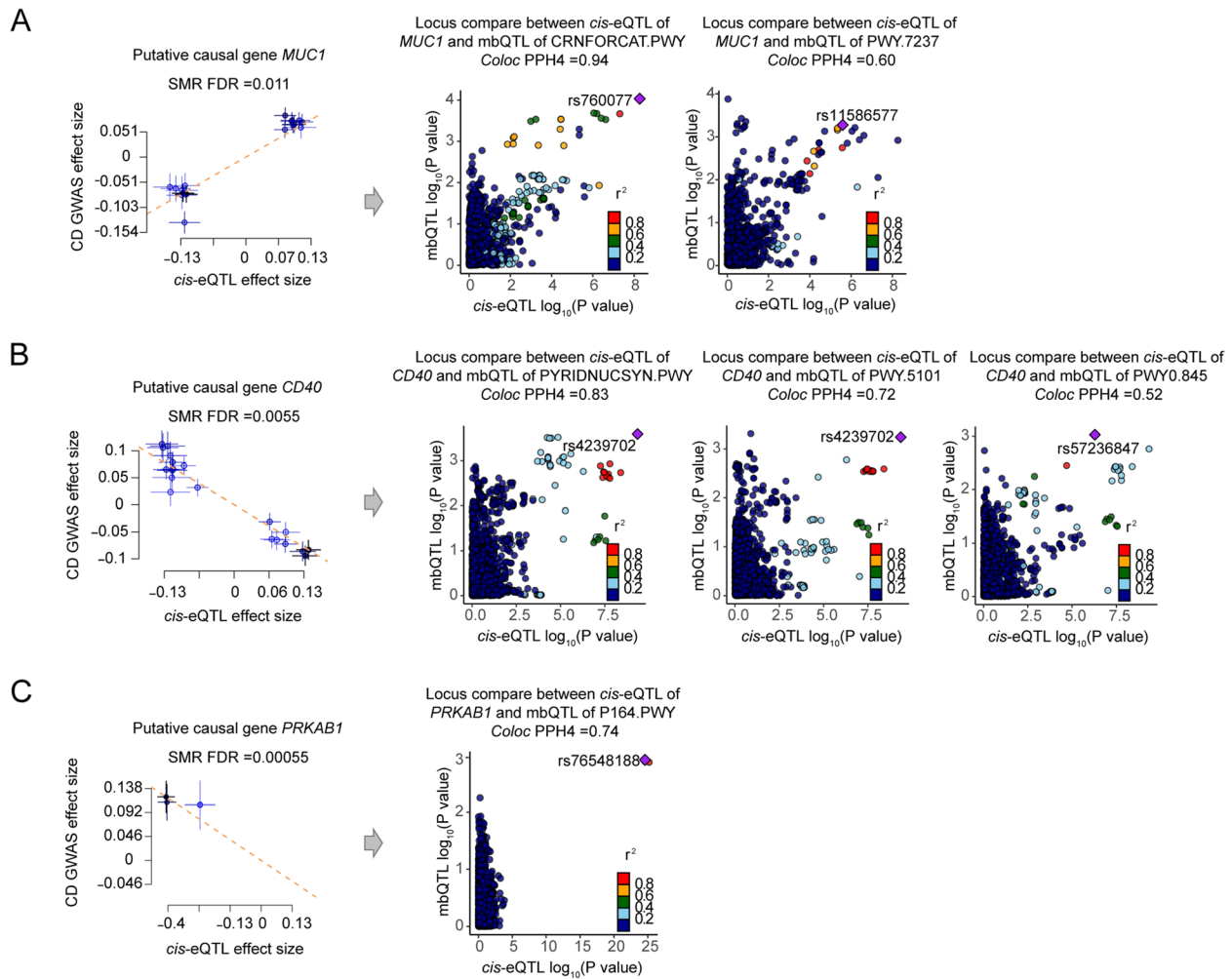Xu *et al. BMC Medicine*   (2023) 21:179

Page 9 of 16

Fig. 4 SMR and colocalization analyses prioritized intestinal causal OS genes and interactions with gut microbial pathways in CD. The left panels indicate the SMR between gene expressions and CD GWAS (all SMR FDR < 0.05; HEIDI test P > 0.05), while the right panels show the locus comparisons between *cis*-eQTLs and mbQTLs by colocalization analysis (all PPH4 > 0.5). The r^2 value indicates the linkage disequilibrium (LD) between the variants and the top SNPs. **A–C** The genes of *MUC1*, *CD40*, and *PRKAB1*, respectively

(PWY0.845; PPH4 = 0.52). These microbiota-derived substances have also been reported to be correlated with intestinal inflammation [61–65]. For instance, nicotinamide adenine dinucleotide can be produced by certain gut bacteria from L-aspartate [66], which is associated with gut inflammation in IBD [63, 64]. L-isoleucine supplementation alleviates intestinal inflammation in vivo and in vitro [65]. Thus, we hypothesize a mechanism by which genetic variants regulate *CD40* expression and interact with inflammation-related microbial activities and therefore contribute to CD pathogenesis.

Another candidate causal gene is *PRKAB1* (Fig. 4C). This study showed that *PRKAB1* expression was also negatively associated with the risk of developing CD (beta$_{SMR}$ = − 0.30). Additionally, its expression shared genetic effects with microbial purine nucleobase

degradation (P164.PWY; PPH4 = 0.74), indicating a potential interaction between gene expression and intestinal microbial nucleotide metabolism.

External replication of OS DEGs and plausible gene–microbiota interactions in the FAH-SYS cohort

To confirm the prioritized genes above, we first validated the DEG results from the meta-analysis using intestinal transcriptomic sequencing data in an independent FAH-SYS cohort (CD n = 46; HC n = 44; the "Methods" section; Additional file 2: Table S12). A lenient significance threshold (P < 0.05) was adopted considering the limited sample size. In total, 437 out of 438 intestinal DEGs from the meta-analysis were tested between patients with CD and HCs. Of note, 388 (88.79%) genes were consistently upregulated or downregulated under disease conditions
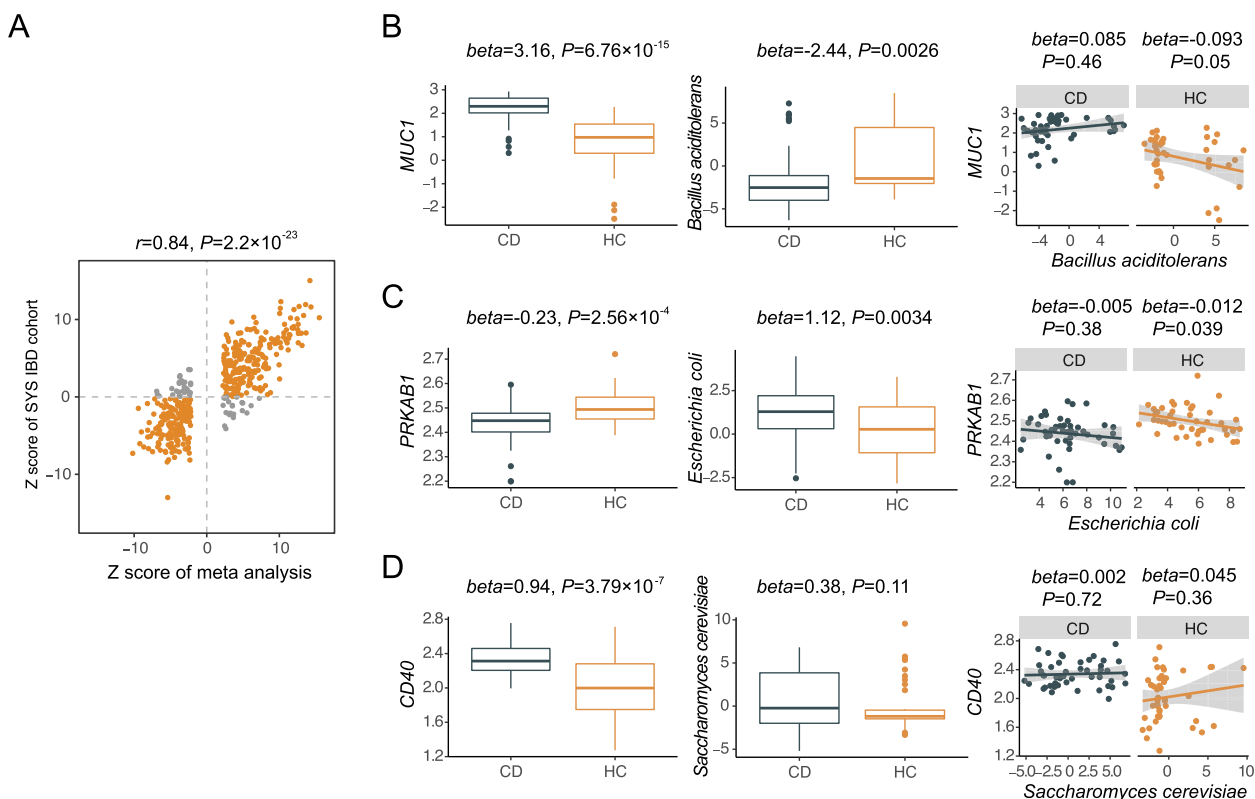
Xu *et al. BMC Medicine*   (2023) 21:179

Page 10 of 16

Fig. 5 External cohort validation. First Affiliated Hospital of Sun Yat-sen University (FAH-SYS) cohort with both intestinal bulk RNA-seq and fecal metagenomics data included in validation analysis. **A** A total of 388 out of 437 (88.79%) differentially expressed genes (DEGs) identified from the meta-analysis show consistent changes between patients with CD and HCs (Spearman rank's correlation coefficient $r_s = 0.84$, $P = 2.2 \times 10^{-23}$). The x- and y-axes indicate the *Z* score estimated from the meta-analysis and FAH-SYS cohort, respectively. Yellow dots represent the 388 consistent DEGs, while gray dots are non-validated. **B**–**D** Validation of DEGs, pathway-related taxa, and the association between taxa and gene expressions, respectively (from left to right). The upper to lower panels are for the *MUC1*, *PRKAB1*, and *CD40* genes, respectively. The taxa probably involved in PWY.7237, P164.PWY, and PWY.5101 were selected from MetaCyc database annotation

(Spearman rank's correlation coefficient $r_s = 0.84$; $P = 2.2 \times 10^{-23}$; Fig. 5A). A total of 320 (73.23%) genes achieved a *P* value < 0.05 (Additional file 2: Table S13), suggesting a robust change in OS gene expression in CD.

Colocalization analysis demonstrated candidate genetic regulations of both intestinal gene expression and gut microbiota; however, little information is available on the correlative directions between the latter two. We then integrated the paired fecal metagenomic and intestinal gene expression data from the FAH-SYS cohort. Two genera and three species potentially involved in colocalized metabolic pathways were evaluated for their associations with intestinal gene expression (the "Methods" section). Five gene–microbiota pairs showed significant correlations in the FAH-SYS cohort (*P* < 0.05; Additional file 2: Table S14). Intestinal *MUC1* gene expression was upregulated in patients with CD compared with that in HCs (beta = 3.16; $P = 6.76 \times 10^{-15}$; Fig. 5B). *Bacillus aciditolerans*, a species from previously reported probiotics involved in myo-inositol degradation (PWY.7237),

was less abundant in the CD group than in the HC group (beta = − 2.44; *P* = 0.0026). A negative correlation between *MUC1* expression and *B. aciditolerans* abundance was observed in the HC group (beta = − 0.093; *P* = 0.05). *E. coli* is a purine utilizer (P164.PWY) and an opportunistic pathogen associated with various immune-related diseases [67]. We identified a higher abundance of *E. coli* in CD while it was associated with a lower level of *PRKAB1* gene expression (beta = − 0.012; *P* = 0.039 in HCs; Fig. 5C). Meanwhile, we observed an increasing trend in *S. cerevisiae* abundance in the FAH-SYS CD group (beta = 0.38; *P* = 0.11), which participates in L-isoleucine biosynthesis (PWY.5101). The association between *S. cerevisiae* and *CD40* expression was also insignificant (Fig. 5D). Notably, the lack of a significant correlation between these intestinal gene expressions and microbiota in CD was likely due to dysbiosis in the disease with disrupted host–microbiota interactions [68]. Nevertheless, the results from an external cohort were consistent with eQTL–mbQTL colocalization. This

Xu *et al. BMC Medicine*   (2023) 21:179

Page 11 of 16

provides complementary evidence of the CD-causal effect of *MUC1* expression and the protective role of *PRKAB1* expression upon interaction with gut microbiota.

Discussion

To the best of our knowledge, this study is the first to leverage a multi-omics integration method to detect putative causal OS genes and the underlying mechanisms in CD using blood and intestinal tissues. We identified 438 OS-related DEGs out of 817 potential genes in CD in a sizable meta-analysis of intestinal transcriptome data and successfully validated these DEGs in our cohort. Integration of GWAS with the eQTLs and mQTLs of these DEGs from the peripheral blood prioritized five putative OS genes and their regulatory elements associated with CD onset: *BAD*, *SHC1*, *STAT3*, *MUC1*, and *GPX3*. Moreover, the integration of intestinal eQTL data also identified five candidate causal genes, of which *MUC1*, *CD40*, and *PRKAB1* were involved in intestinal gene–microbiota interactions through further colocalization analysis. Differentially expressed OS genes in CD can either be a cause or a collateral effect of intestinal inflammation; our study is therefore fundamental as an attempt to fill the gaps in our understanding of discriminating between either causally or remotely involved OS genes and pinpointing the relevant interactions in CD in a genomic context.

Recently, more and more blood-based biomarkers are being used to diagnose, monitor, and predict IBD activities, and blood tissue may serve as a valuable proxy in terms of characterizing genetic effects on gene expression and understanding the complex etiology of IBD. Using an SMR analysis with blood tissue, we detected five putative causal associations between OS genes and CD susceptibility through genetically epigenomic and transcriptomic regulation, suggesting a vital role of epigenetic factors and gene expression in the disease onset. Among these genes (*BAD*, *SHC1*, *STAT3*, *MUC1*, and *GPX3*), the causal roles of *STAT3* and *MUC1* have been extensively characterized in CD [69–71]. For instance, *STAT3* plays a crucial role in many cellular processes, including cell growth and apoptosis in response to cellular stimuli, and has been regarded as a CD susceptibility gene according to previous GWASs [72–74]. Moreover, a T cell-specific *STAT3* deletion has been reported to ameliorate dextran sulfate sodium-induced colitis in mice by reducing the inflammatory response [75]. Importantly, our study confirmed that an increased transcript level of *STAT3* may lead to an increased CD risk (beta$_{SMR}$=0.70). Additionally, we revealed that DNAm in enhancer regions negatively regulated *STAT3* expression, suggesting a link between DNAm, *STAT3* expression, and CD risk. More importantly, another three candidate genes lacking

intensive study were identified from blood tissue that might be causal to CD: *GPX3*, *SHC1*, and *BAD*. *GPX3* is involved in the redox-sensitive KEAP1-NRF2/ARE signaling system which is considered a pivotal target in maintaining cellular homeostasis under OS, inflammatory conditions, and pro-apoptotic conditions [76]. To date, studies on the *GPX3* gene, which is a target gene of NRF2, have mainly focused on cancers, including colitis-associated carcinoma; for example, *Gpx3*-deficient mice exhibited increased tumor number and inflammation, suggesting a protective role of *GPX3* in colitis-associated carcinoma [77]. Similarly, our findings indicated a negative (protective) effect of *GPX3* expression on CD susceptibility (beta$_{SMR}$= −0.15). *SHC1* is a signaling adapter molecule that is heavily understudied in CD. This gene encodes three main isoforms, and the most extended isoform (p66Shc) is a central regulator of OS in mitochondria and cells across multiple diseases [78–80]. This work showed that four DNAm sites near the promoter regions were significantly associated with *SHC1* and CD, indicating a co-regulatory pattern involving multiple epigenetic regulatory elements [81]. BAD protein is a key participant in mitochondria-dependent apoptosis and pathophysiological processes that involve the regulation of OS [82]. Although previous CD GWAS data have identified genetic variants located nearby OS genes like *SHC1* and *BAD* [83], it remains unclear whether these genes have a causal effect on CD. Based on our SMR analysis, we hypothesize that genetic variants could regulate the expression of these genes through DNA methylation, thereby affecting CD pathogenesis.

Tissue- and cell-specific gene expression has been shown to elucidate different biological molecular mechanisms [55, 84]. CD is a gastrointestinal disease, and studying its genetic effects on OS gene expression in the intestine using intestinal eQTLs (the most pertinent tissue type) may be more meaningful than that in the blood. As the intestinal barrier directly contacts with luminal microbes and oxidized compounds from external environment factors and senses the recurrent oxidative changes, OS genes in the intestine may be associated with CD through host–microbiota interactions. Our SMR-based analysis pinpointed *MUC1*, *CD40*, *PARK7*, *PRKAB1*, and *NDUFS1* as putative causal genes in intestinal tissue, of which *MUC1* was also of interest in the blood. However, the association between *MUC1* expression and CD differed in the blood compared to that in the intestine, suggesting tissue-specific effects during the onset of CD. Moreover, we identified novel genes in this context that might contribute to CD pathogenesis, such as *PRKAB1* and *NDUFS1*. Three genes, *MUC1*, *CD40*, and *PRKAB1*, were further prioritized when considering the interactions between host genetics and microbiota.

Xu *et al. BMC Medicine*   (2023) 21:179

Page 12 of 16

MUC1 encodes a vital constituent of mucus and is over-expressed and hypo-glycosylated in the development of inflammation and IBD given its role in regulating intestinal barrier function upon multiple stimuli, including OS [71, 85–87]. Furthermore, *Muc1* knockout mice are resistant to dextran sulfate sodium-induced acute intestinal injury [70]. This is consistent with our findings which confirmed that high expression of *MUC1* increases the risk of developing CD. Additionally, our study colocalized the genetic regulations of *MUC1* expression and gut microbiota. Microbial creatinine degradation and myo-inositol degradation shared genetic effects with *MUC1* expression, suggesting the potential interactions between the gene and microbiota. Creatinine supplementation is identified as a potential therapeutic treatment for IBD; creatinine can be degraded to creatine by gut microbiota [88]. Creatinine clearance is associated with reduced inflammation and decreased fibrosis [89]. In addition, myo-inositol derived from dietary phytate can be converted to short-chain fatty acids through colonic bacterial phytase activity [90]. *B. aciditolerans*, one of the predicted pathway-related taxa, was negatively correlated with *MUC1* expression in the FAH-SYS cohort. This suggests that high *MUC1* expression accompanied by decreased beneficial microbial activities could confer an increased risk of CD. Our study also inferred that *PRKAB1*, which encodes the regulatory subunit of AMP-activated protein kinase that monitors cellular energy status and responds to ROS [91], was a CD-protective OS gene. We suggest that genetic regulation of *PRKAB1* expression is associated with CD onset (beta$_{SMR}$ = − 0.30). Moreover, *E. coli* and related purine nucleobase degradation pathways may interact with host *PRKAB1* expression. IBD-associated *E. coli* strains have been reported to facilitate IBD flares [92]. A negative correlation between *PRKAB1* expression and *E. coli* was consistently observed in the FAH-SYS cohort. Interestingly, a recent study reported the role of *PRKAB1* agonists as barrier-protective therapeutic agents in IBD [93]. However, further evidence based on genetic background (such as knockout mouse models) is needed to precisely explain the potential role of *PRKAB1* in CD.

Integrating multi-omics from multiple tissues enables researchers to dissect GWAS signals, such as the prioritization of genes and disease mechanisms. Peripheral blood tissue has a less direct and significant effect on CD than intestinal tissue. However, its significance in generating epigenomic, transcriptomic, and proteomic evidence for identifying causally involved genes and therapeutically relevant targets is well recognized [19, 94, 95]. We prioritized a list of novel genes and DNAm sites for follow-up functional studies using the largest up-to-date CD GWAS and OS-targeted approach. More importantly, this is the first study providing evidence to support a causal role of OS genes interacting with the gut microbiota in intestinal tissue. Despite the moderate associations between host genetics and gut microbiota [46], we observed common genetic regulations of intestinal gene expression and bacterial metabolic potentials. Different bacteria harboring shared genomic contents can participate in the same metabolic functions [96, 97]. However, no individual taxa were significantly colocalized with the intestinal gene expression in the current analysis. This is likely owing to the low statistical power to detect zero-inflated taxa data in mbQTL studies [44]. Nevertheless, we used an external multi-omics cohort to confirm the association between the expression of these genes and pathway-related bacterial abundance. In addition, microbiota detected from fecal and intestinal tissues showed considerable differences [98–100], which might explain the small effect sizes from the intestinal gene–fecal microbiota associations.

Some limitations of this study warrant recognition. First, the meta-analysis of intestinal DEGs included different data resources (microarray and bulk RNA-seq with varying sample sizes) which could impose heterogeneity. However, we successfully replicated over 80% of the DEGs in an independent cohort with pronounced transcriptomic alterations of the OS gene family in patients with CD compared with the controls. Second, cell type-dependent eQTLs vary with disease progression [101, 102]. The eQTLs from the bulk RNA-seq limited the identification of key molecular mechanisms at the intestinal cell level (enterocytes, immunocytes, fibrocytes) related to CD. Third, we only focused on the *cis*-regions for OS genes in the analysis, despite the possibility that *trans*-eQTL SNPs (SNP and the center of the gene > 5 Mb) may have a widespread impact on regulatory networks [41]. Fourth, we used a Bayesian colocalization method which relies on an assumption that two straits share the same single genomic variant while the case of multiple causal variants is under-explored [103]. Finally, functional experiments are still needed to validate our findings. Moreover, as multiple factors can influence the expression of OS genes, we believe that integrating other omics data at different molecular levels (such as those of proteins and metabolites) with large sample sizes may lead to novel discoveries and improve the characterization of putatively involved causal mechanisms of OS in CD.

Conclusions
This study expands our knowledge of the potential causality of OS and the underlying biological mechanisms in CD based on a multi-omics MR approach. We demonstrated that CD onset putatively results from a number

Xu *et al. BMC Medicine*   (2023) 21:179

Page 13 of 16

of candidate OS genes through DNAm, gene expression, and interaction with gut microbiota. Host–microbiota interactions between our newly identified causal OS genes and microbial taxa and pathways are worth studying at a functional level to gain more in-depth insights into the underlying biological mechanisms. This study advances fundamental research into the role of OS in CD and pinpoints potentially novel therapeutic targets for clinical practice.

Abbreviations

BH	Benjamini–Hochberg
CD	Crohn's disease
CSEA	Cell type-specific expression analysis
CSEA-DB	Cell type-Specific Enrichment Analysis DataBase
DEG	Differentially expressed gene
DMP	Dutch Microbiome Project
DNAm	DNA methylation
eQTL	Expression quantitative trait locus
FAH-SYS	First Affiliated Hospital of Sun Yat-sen University
FDR	False discovery rate
GEO	Gene Expression Omnibus
GTEx	Genotype-Tissue Expression
GWAS	Genome-wide association study
HC	Healthy control
HEIDI	Heterogeneity in dependent instruments
IBD	Inflammatory bowel disease
IVW	Inverse variance weighting
LD	Linkage disequilibrium
mbQTL	Fecal microbial QTL
MeCS	Meta-analysis of *cis*-eQTL in correlated samples
mQTL	DNA methylation QTL
MR	Mendelian randomization
OS	Oxidative stress
RNA-seq	RNA sequencing
ROS	Reactive oxygen species
SMR	Summary data-based Mendelian randomization
SNP	Single nucleotide polymorphism
TSS	Transcription start site

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12916-023-02878-8.

Additional file 1. Supplementary Methods.

Additional file 2: Table S1. Characteristics of six transcriptome datasets used in the study. **Table S2.** 817 OS-related genes with a relevance score ≥ 7 obtained from GeneCards. **Table S3.** Meta-analysis of 708 differentially expressed OS genes from six datasets. **Table S4.** Cell type-specific expression analysis (CSEA) of 438 intestinal OS DEGs. **Table S5.** Summary-based Mendelian randomization (SMR) analysis from blood gene expression to CD (FDR < 0.05). **Table S6.** SMR analysis from blood DNA methylation to CD (FDR < 0.05). **Table S7.** SMR analysis from blood DNA methylation to gene expression (FDR < 0.05). **Table S8.** Regulatory component annotation of 665 DNA methylation sites. **Table S9.** Meta-analysis between GTEx and 1000IBD intestinal cis-eQTLs (FDR < 0.05). **Table S10.** SMR analysis from intestinal gene expression to CD (FDR < 0.05). **Table S11.** Colocalization analysis between gene expression and gut microbiota (PPH4 > 0.5). **Table S12.** Demographic characteristics of FAH-SYS IBD multi-omics cohort. **Table S13.** External cohort validation: intestinal DEG results between CD and HC. **Table S14.** External cohort validation: associations between intestinal gene expression and gut microbiota.

Additional file 3: Fig. S1. Regulatory component annotation of 665 DNA methylation (DNAm) sites. A) DNAm sites annotated with active or inactive chromatin states were enriched in different blood cell types. B) DNAm sites annotated with histone markers were enriched in different blood cell types. Reference of active chromatin states: active transcription start site (TSS)-proximal promoter states (TssA, TssAFlnk), a transcribed state at the 5′ and 3′ ends of genes showing both promoter and enhancer signatures (TxFlnk), actively transcribed states (Tx, TxWk), enhancer states (Enh, EnhG), and a state associated with zinc finger protein genes (ZNF/Rpts). Reference of inactive chromatin states: constitutive heterochromatin (Het), bivalent regulatory states (TssBiv, BivFlnk, EnhBiv), repressed polycomb states (ReprPC, ReprPCWk), and quiescent state (Quies). Reference histone marker: H3K4me1 is enriched at active and primed enhancers; H3K4me3 is a modification that is associated with transcriptionally active/poised chromatin.

Additional file 4: Fig. S2. Manhattan plots of colocalization between intestinal *cis*-eQTLs and mbQTLs. Six pairs of *cis*-eQTLs and mbQTLs with PPH4 > 0.5 were plotted for illustration: *MUC1*–CARNFORCAT.PWY, *MUC1*–PWWY.7237, *CD40*–PYRIDNUCSYN.PWY, *CD40*–PWY.5101, *CD40*–PWY.0845 and *PRKAB1*–P164.PWY. All microbial pathways were annotated by MetaCyc database (https://metacyc.org/). The *x*-axis shows the chromosomal positions while the *y*-axis indicates the –log10 *P* values of SNPs.

Authors' contributions

SH, RF, and MC conceptualized, designed, and supervised the study. SX, SH, and XL collected the data and performed the data analysis. SX, SH, RF, and XL wrote and revised the manuscript. SZ, RM, and LX were responsible for recruiting patients, collecting the samples, and recording clinical data. CQ, ZZ1, LC, ARB, and ML provided advice for the study and revised the manuscript for important intellectual contents. YZ, CT, YH, and ZZ2 revised the manuscript. The authors read and approved the final manuscript.

Authors' information

Twitter handle: @Shixian00000 (Shixian Hu).

Availability of data and materials

The standard tools used in this study (including SMR and MeCS) are available at https://yanglab.westlake.edu.cn/software/smr. The pipelines and scripts were stored at https://github.com/EMO-Consortium/GeneticCorrelation. Six publicly available transcriptome datasets were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) [30, 32, 34, 35, 37, 39]. GWAS summary statistics for CD were downloaded from ftp://ftp.sanger.ac.uk/pub/project/humgen/summary_statistics/human/2016-11-07/. Publicly available summary statistics of blood eQTLs and mQTLs were obtained from http://www.eqtlgen.org and https://yanglab.westlake.edu.cn/software/smr/#mQTLsummarydata, respectively. Intestinal eQTL data were downloaded from https://gtexportal.org/home/datasets and https://ega-archive.org/studies/EGAS00001002702 under EGAD00001006789. mbQTL data from the DMP study can be found at https://dutchmicrobiomeproject.molgeniscloud.org. Raw metagenomic data of the FAH-SYS cohort were deposited in the NCBI public repository (Bioproject #PRJNA793776). Raw intestinal RNA-seq data are available from the corresponding authors upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the Medical Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University (Application ID: [2016] 113). All participants included in the FAH-SYS cohort provided informed consent.

Xu *et al. BMC Medicine*   (2023) 21:179

Page 14 of 16

Author details
[1]Department of Gastroenterology, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, Guangdong, China. [2]Microbiome Medicine Center, Division of Laboratory Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong, China. [3]Department of Computational Biology for Individualised Medicine, Centre for Individualised Infection Medicine & TWINCORE, Joint Ventures Between the Helmholtz Centre for Infection Research and the Hannover Medical School, Hannover, Germany. [4]Changzhou Medical Center, The Affiliated Changzhou No. 2 People's Hospital of Nanjing Medical University, Nanjing Medical University, Changzhou, Jiangsu, China. [5]Department of Cardiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing Medical University, Nanjing, Jiangsu, China. [6]Institute of Precision Medicine, The First Affiliated Hospital, Sun Yat-Sen University, Guangzhou, Guangdong, China. [7]Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. [8]Zhongshan School of Medicine, Center for Precision Medicine, Sun Yat-Sen University, Guangzhou, Guangdong, China. [9]Department of Gastroenterology, Guangxi Hospital Division of The First Affiliated Hospital, Sun Yat-Sen University, Nanning, Guangxi, China.

References
1. Baumgart DC, Sandborn WJ. Crohn's disease. Lancet. 2012;380(9853):1590–605.
2. Bernstein CN. Treatment of IBD: where we are and where we are going. Am J Gastroenterol. 2015;110(1):114–26.
3. Sies H. Oxidative stress: a concept in redox biology and medicine. Redox Biol. 2015;4:180–3.
4. Singer II, Kawka DW, Schloemann S, Tessner T, Riehl T, Stenson WF. Cyclooxygenase 2 is induced in colonic epithelial cells in inflammatory bowel disease. Gastroenterology. 1998;115(2):297–306.
5. Meriwether D, Sulaiman D, Volpe C, Dorfman A, Grijalva V, Dorreh N, et al. Apolipoprotein A-I mimetics mitigate intestinal inflammation in COX2-dependent inflammatory bowel disease model. J Clin Invest. 2019;129(9):3670–85.
6. Martín MC, Martinez A, Mendoza JL, Taxonera C, Díaz-Rubio M, Fernández-Arquero M, et al. Influence of the inducible nitric oxide synthase gene (NOS2A) on inflammatory bowel disease susceptibility. Immunogenetics. 2007;59(11):833–7.
7. Grasberger H, Gao J, Nagao-Kitamoto H, Kitamoto S, Zhang M, Kamada N, et al. Increased expression of Duox2 is an epithelial response to mucosal dysbiosis required for immune homeostasis in mouse intestine. Gastroenterology. 2015;149(7):1849–59.
8. Makhezer N, Ben Khemis M, Liu D, Khichane Y, Marzaioli V, Tlili A, et al. NOX1-derived ROS drive the expression of lipocalin-2 in colonic epithelial cells in inflammatory conditions. Mucosal Immunol. 2019;12(1):117–31.
9. Dang PM, Rolas L, El-Benna J. The dual role of reactive oxygen species-generating nicotinamide adenine dinucleotide phosphate oxidases in gastrointestinal inflammation and therapeutic perspectives. Antioxid Redox Signal. 2020;33(5):354–73.
10. Alemany-Cosme E, Sáez-González E, Moret I, Mateos B, Iborra M, Nos P, et al. Oxidative stress in the pathogenesis of Crohn's disease and the interconnection with immunological response, microbiota, external environmental factors, and epigenetics. Antioxidants (Basel). 2021;10(1):64.
11. Ventham NT, Kennedy NA, Nimmo ER, Satsangi J. Beyond gene discovery in inflammatory bowel disease: the emerging role of epigenetics. Gastroenterology. 2013;145(2):293–308.
12. Yara S, Lavoie JC, Levy E. Oxidative stress and DNA methylation regulation in the metabolic syndrome. Epigenomics. 2015;7(2):283–300.
13. Kang Q, Yang C. Oxidative stress and diabetic retinopathy: molecular mechanisms, pathogenetic role and therapeutic implications. Redox Biol. 2020;37:101799.
14. Guo Y, Yu S, Zhang C, Kong AN. Epigenetic regulation of Keap1-Nrf2 signaling. Free Radic Biol Med. 2015;88(Pt B):337–49.
15. Krzystek-Korpacka M, Kempiński R, Bromke MA, Neubauer K. Oxidative stress markers in inflammatory bowel diseases: systematic review. Diagnostics (Basel). 2020;10(8):601.
16. Sazonovs A, Stevens CR, Venkataraman GR, Yuan K, Avila B, Abreu MT, et al. Large-scale sequencing identifies multiple genes and rare variants associated with Crohn's disease susceptibility. Nat Genet. 2022;54(9):1275–83.
17. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005–12.
18. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015;518(7539):337–43.
19. Wu Y, Zeng J, Zhang F, Zhu Z, Qi T, Zheng Z, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. Nat Commun. 2018;9(1):918.
20. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 2016;48(5):481–7.
21. Kapoor M, Chao MJ, Johnson EC, Novikova G, Lai D, Meyers JL, et al. Multi-omics integration analysis identifies novel genes for alcoholism with potential overlap with neurodegenerative diseases. Nat Commun. 2021;12(1):5071.
22. Bourgonje AR, Feelisch M, Faber KN, Pasch A, Dijkstra G, van Goor H. Oxidative stress and redox-modulating therapeutics in inflammatory bowel disease. Trends Mol Med. 2020;26(11):1034–46.
23. Hannon E, Weedon M, Bray N, O'Donovan M, Mill J. Pleiotropic effects of trait-associated genetic variation on dna methylation: utility for refining gwas loci. Am J Hum Genet. 2017;100(6):954–9.
24. Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet. 2021;53(2):156–65.
25. Di'Narzo AF, Houten SM, Kosoy R, Huang R, Vaz FM, Hou R, et al. Integrative analysis of the inflammatory bowel disease serum metabolome improves our understanding of genetic etiology and points to novel putative therapeutic targets. Gastroenterology. 2022;162(3):828-43.e11.
26. Qiu X, Hou QH, Shi QY, Jiang HX, Qin SY. Identification of hub prognosis-associated oxidative stress genes in pancreatic cancer using integrated bioinformatics analysis. Front Genet. 2020;11:595361.
27. Fan J, Cao S, Chen M, Yao Q, Zhang X, Du S, et al. Investigating the AC079305/DUSP1 axis as oxidative stress-related signatures and immune infiltration characteristics in ischemic stroke. Oxid Med Cell Longev. 2022;2022:8432352.
28. Sun X, Huang X, Sun X, Chen S, Zhang Z, Yu Y, et al. Oxidative stress-related lncRNAs are potential biomarkers for predicting prognosis and immune responses in patients with LUAD. Front Genet. 2022;13:909797.
29. Vancamelbeke M, Vanuytsel T, Farré R, Verstockt S, Ferrante M, Van Assche G, et al. Genetic and transcriptomic bases of intestinal epithelial barrier dysfunction in inflammatory bowel disease. Inflamm Bowel Dis. 2017;23(10):1718–29.
30. Arijs I, Vancamelbeke M, Vanhove W, De Hertogh G, Schuit F, Rutgeerts P, et al. Mucosal gene expression profiling in patients with inflammatory bowel disease study. GEO https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE75214. 2017.
31. VanDussen KL, Stojmirović A, Li K, Liu TC, Kimes PK, Muegge BD, et al. Abnormal small intestinal epithelial microvilli in patients with Crohn's disease. Gastroenterology. 2018;155(3):815–28.
32. VanDussen KL, Stojmirović A, Li K, Liu T, Kimes PK, Muegge BD, et a. Intestinal epithelial microvilli are abnormal in Crohn's disease. GEO https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112366. 2019.
33. Keir ME, Fuh F, Ichikawa R, Acres M, Hackney JA, Hulme G, et al. Regulation and role of αE integrin and gut homing integrins in migration and

Xu *et al. BMC Medicine*   (2023) 21:179

Page 15 of 16

retention of intestinal lymphocytes during inflammatory bowel disease. J Immunol. 2021;207(9):2245–54.

34. Hackney JA, Keir ME. Embark cross-sectional study of inactive and active lesions from the ileum and colon of inflammatory bowel disease patients. GEO https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE179285. 2021.

35. D'alessio S, Ungaro F, Massimino L, Lamparelli LA, Danese S. Characterization of the IL23-IL17 immune axis in IBD patients. GEO https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165512. 2021.

36. Mo A, Krishnakumar C, Arafat D, Dhere T, Iskandar H, Dodd A, et al. African ancestry proportion influences ileal gene expression in inflammatory bowel disease. Cell Mol Gastroenterol Hepatol. 2020;10(1):203–5.

37. Mo A, Arafat D, Dodd A, Prince J, Kugathasan S, Gibson G. Ileum transcriptomic profiling in inflammatory bowel disease. GEO https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137344. 2020.

38. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569(7758):655–62.

39. Arze C, Plichata D, Bishai J. Longitudinal multi-omics of the human microbiome in inflammatory bowel disease. GEO https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111889. 2018.

40. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet. 2017;49(2):256–61.

41. Võsa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet. 2021;53(9):1300–10.

42. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318–30.

43. Hu S, Uniken Venema WT, Westra HJ, Vich Vila A, Barbieri R, Voskuil MD, et al. Inflammation status modulates the effect of host genetic variation on intestinal gene expression in inflammatory bowel disease. Nat Commun. 2021;12(1):1122.

44. Lopera-Maya EA, Kurilshikov A, van der Graaf A, Hu S, Andreu-Sánchez S, Chen L, et al. Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. Nat Genet. 2022;54(2):143–51.

45. Tian Z, Zhuang X, Zhuo S, Zhu Y, Hu S, Zhao M, et al. Dietary inflammatory potential mediated gut microbiota and metabolite alterations in Crohn's disease: A fire-new perspective. Clin Nutr. 2022;41(6):1260–71.

46. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018;555(7695):210–5.

47. Gacesa R, Kurilshikov A, Vich Vila A, Sinha T, Klaassen MAY, Bolte LA, et al. Environmental factors shaping the gut microbiome in a Dutch population. Nature. 2022;604(7907):732–9.

48. Huang Y, Shan Y, Zhang W, Lee AM, Li F, Stranger BE, Huang RS, et al. Deciphering genetic causes for sex differences in human health through drug metabolism and transporter genes. Nat Commun. 2023;14(1):175.

49. Arvanitis M, Tayeb K, Strober BJ, Battle A. Redefining tissue specificity of genetic regulation of gene expression in the presence of allelic heterogeneity. Am J Hum Genet. 2022;109(2):223–39.

50. Jukic A, Bakiri L, Wagner EF, Tilg H, Adolph TE. Calprotectin: from biomarker to biological function. Gut. 2021;70(10):1978–88.

51. Bourgonje AR, Alexdottir MS, Otten AT, Loveikyte R, Bay-Jensen AC, Pehrsson M, et al. Serological biomarkers of type I, III and IV collagen turnover are associated with the presence and future progression of stricturing and penetrating Crohn's disease. Aliment Pharmacol Ther. 2022;56(4):675–93.

52. Burgueño JF, Fritsch J, González EE, Landau KS, Santander AM, Fernández I, et al. Epithelial TLR4 signaling activates DUOX2 to induce microbiota-driven tumorigenesis. Gastroenterology. 2021;160(3):797-808.e6.

53. Butturini E, Carcereri de Prati A, Mariotto S. Redox regulation of STAT1 and STAT3 signaling. Int J Mol Sci. 2020;21(19):7034.

54. Nirgude S, Choudhary B. Insights into the role of GPX3, a highly efficient plasma antioxidant, in cancer. Biochem Pharmacol. 2021;184:114365.

55. Kabakchiev B, Silverberg MS. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. Gastroenterology. 2013;144(7):1488-96.e3.

56. Hu S, Vich Vila A, Gacesa R, Collij V, Stevens C, Fu JM, et al. Whole exome sequencing analyses reveal gene-microbiota interactions in the context of IBD. Gut. 2021;70(2):285–96.

57. Lee JR, Koretzky GA. Production of reactive oxygen intermediates following CD40 ligation correlates with c-Jun N-terminal kinase activation and IL-6 secretion in murine B lymphocytes. Eur J Immunol. 1998;28(12):4188–97.

58. Urbich C, Dernbach E, Aicher A, Zeiher AM, Dimmeler S. CD40 ligand inhibits endothelial cell migration by increasing production of endothelial reactive oxygen species. Circulation. 2002;106(8):981–6.

59. Rizvi M, Pathak D, Freedman JE, Chakrabarti S. CD40-CD40 ligand interactions in oxidative stress, inflammation and vascular disease. Trends Mol Med. 2008;14(12):530–8.

60. Jacobs BM, Taylor T, Awad A, Baker D, Giovanonni G, Noyce AJ, et al. Summary-data-based Mendelian randomization prioritizes potential druggable targets for multiple sclerosis. Brain Commun. 2020;2(2):fcaa119.

61. Benight NM, Stoll B, Chacko S, da Silva VR, Marini JC, Gregory JF, et al. B-vitamin deficiency is protective against DSS-induced colitis in mice. Am J Physiol Gastrointest Liver Physiol. 2011;301(2):G249–59.

62. Chiocchetti A, Prodam F, Dianzani U. Homocysteine and folate in inflammatory bowel disease: can reducing sulfur reduce suffering? Dig Dis Sci. 2018;63(12):3161–3.

63. Gerner RR, Klepsch V, Macheiner S, Arnhard K, Adolph TE, Grander C, et al. NAD metabolism fuels human and mouse intestinal inflammation. Gut. 2018;67(10):1813–23.

64. Ning L, Shan G, Sun Z, Zhang F, Xu C, Lou X, et al. Quantitative proteomic analysis reveals the deregulation of nicotinamide adenine dinucleotide metabolism and CD38 in inflammatory bowel disease. Biomed Res Int. 2019;2019:3950628.

65. Mao X, Sun R, Wang Q, Chen D, Yu B, He J, et al. l-Isoleucine administration alleviates DSS-induced colitis by regulating TLR4/MyD88/NF-κB pathway in rats. Front Immunol. 2021;12:817583.

66. Begley TP, Kinsland C, Mehl RA, Osterman A, Dorrestein P. The biosynthesis of nicotinamide adenine dinucleotides in bacteria. Vitam Horm. 2001;61:103–19.

67. Denamur E, Clermont O, Bonacorsi S, Gordon D. The population genetics of pathogenic Escherichia coli. Nat Rev Microbiol. 2021;19(1):37–54.

68. Priya S, Burns MB, Ward T, Mars RAT, Adamowicz B, Lock EF, et al. Identification of shared and disease-specific host gene-microbiome associations across human diseases using multi-omic integration. Nat Microbiol. 2022;7(6):780–95.

69. Li Y, de Haar C, Peppelenbosch MP, van der Woude CJ. New insights into the role of STAT3 in IBD. Inflamm Bowel Dis. 2012;18(6):1177–83.

70. Nishida A, Lau CW, Zhang M, Andoh A, Shi HN, Mizoguchi E, et al. The membrane-bound mucin Muc1 regulates T helper 17-cell responses and colitis in mice. Gastroenterology. 2012;142(4):865-74.e2.

71. Breugelmans T, Van Spaendonk H, De Man JG, De Schepper HU, Jauregui-Amezaga A, Macken E, et al. In-depth study of transmembrane mucins in association with intestinal barrier dysfunction during the course of T cell transfer and DSS-induced colitis. J Crohns Colitis. 2020;14(7):974–94.

72. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet. 2008;40(8):955–62.

73. Sato K, Shiota M, Fukuda S, Iwamoto E, Machida H, Inamine T, et al. Strong evidence of a combination polymorphism of the tyrosine kinase 2 gene and the signal transducer and activator of transcription 3 gene as a DNA-based biomarker for susceptibility to Crohn's disease in the Japanese population. J Clin Immunol. 2009;29(6):815–25.

74. Ferguson LR, Han DY, Fraser AG, Huebner C, Lam WJ, Morgan AR, et al. Genetic factors in chronic inflammation: single nucleotide polymorphisms in the STAT-JAK pathway, susceptibility to DNA damage and Crohn's disease in a New Zealand population. Mutat Res. 2010;690(1–2):108–15.

75. Kwon SH, Seo EB, Lee SH, Cho CH, Kim SJ, Kim SJ, et al. T cell-specific knockout of STAT3 ameliorates dextran sulfate sodium-induced colitis by reducing the inflammatory response. Immune Netw. 2018;18(4): e30.

Xu *et al. BMC Medicine*   (2023) 21:179

Page 16 of 16

76. Tu W, Wang H, Li S, Liu Q, Sha H. The anti-inflammatory and anti-oxidant mechanisms of the Keap1/Nrf2/ARE signaling pathway in chronic diseases. Aging Dis. 2019;10(3):637–51.

77. Barrett CW, Ning W, Chen X, Smith JJ, Washington MK, Hill KE, et al. Tumor suppressor function of the plasma glutathione peroxidase gpx3 in colitis-associated carcinoma. Cancer Res. 2013;73(3):1245–55.

78. De Marchi E, Baldassari F, Bononi A, Wieckowski MR, Pinton P. Oxidative stress in cardiovascular diseases and obesity: role of p66Shc and protein kinase C. Oxid Med Cell Longev. 2013;2013:564961.

79. Purdom S, Chen QM. p66(Shc): at the crossroad of oxidative stress and the genetics of aging. Trends Mol Med. 2003;9(5):206–10.

80. Wang D, Wang T, Wang R, Zhang X, Wang L, Xiang Z, et al. Suppression of p66Shc prevents hyperandrogenism-induced ovarian oxidative stress and fibrosis. J Transl Med. 2020;18(1):84.

81. Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol. 2013;14(3):R21.

82. Bergmann A. Survival signaling goes BAD. Dev Cell. 2002;3(5):607–8.

83. Yang SK, Hong M, Oh H, Low HQ, Jung S, Ahn S, et al. Identification of loci at 1q21 and 16q23 that affect susceptibility to inflammatory bowel disease in Koreans. Gastroenterology. 2016;151(6):1096-9.e4.

84. Kundu K, Tardaguila M, Mann AL, Watt S, Ponstingl H, Vasquez L, et al. Genetic associations at regulatory phenotypes improve fine-mapping of causal variants for 12 immune-mediated diseases. Nat Genet. 2022;54(3):251–62.

85. Hashash JG, Beatty PL, Critelli K, Hartman DJ, Regueiro M, Tamim H, et al. Altered expression of the epithelial mucin MUC1 accompanies endoscopic recurrence of postoperative Crohn's disease. J Clin Gastroenterol. 2021;55(2):127–33.

86. Paone P, Cani PD. Mucus barrier, mucins and gut microbiota: the expected slimy partners? Gut. 2020;69(12):2232–43.

87. Kadayakkara DK, Beatty PL, Turner MS, Janjic JM, Ahrens ET, Finn OJ. Inflammation driven by overexpression of the hypoglycosylated abnormal mucin 1 (MUC1) links inflammatory bowel disease and pancreatitis. Pancreas. 2010;39(4):510–5.

88. Wallimann T, Hall CHT, Colgan SP, Glover LE. Creatine supplementation for patients with inflammatory bowel diseases: a scientific rationale for a clinical trial. Nutrients. 2021;13(5):1429.

89. Vaziri ND, Liu SM, Lau WL, Khazaeli M, Nazertehrani S, Farzaneh SH, et al. High amylose resistant starch diet ameliorates oxidative stress, inflammation, and progression of chronic kidney disease. PLoS ONE. 2014;9(12):e114881.

90. Bui TPN, Mannerås-Holm L, Puschmann R, Wu H, Troise AD, Nijsse B, et al. Conversion of dietary inositol into propionate and acetate by commensal Anaerostipes associates with host health. Nat Commun. 2021;12(1):4798.

91. Xiao B, Sanders MJ, Carmena D, Bright NJ, Haire LF, Underwood E, et al. Structural basis of AMPK regulation by small molecule activators. Nat Commun. 2013;4:3017.

92. Mirsepasi-Lauridsen HC, Vallance BA, Krogfelt KA, Petersen AM. Escherichia coli pathobionts associated with inflammatory bowel disease. Clin Microbiol Rev. 2019;32(2):e00060-e118.

93. Sahoo D, Swanson L, Sayed IM, Katkar GD, Ibeawuchi SR, Mittal Y, et al. Artificial intelligence guided discovery of a barrier-protective therapy in inflammatory bowel disease. Nat Commun. 2021;12(1):4246.

94. Yao C, Chen G, Song C, Keefe J, Mendelson M, Huan T, et al. Genomewide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nat Commun. 2018;9(1):3268.

95. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman ÅK, Schork A, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. Nat Metab. 2020;2(10):1135–48.

96. Gowda K, Ping D, Mani M, Kuehn S. Genomic structure predicts metabolite dynamics in microbial communities. Cell. 2022;185(3):530-46.e25.

97. Wang Y, Dong Q, Hu S, Zou H, Wu T, Shi J, et al. Decoding microbial genomes to understand their functional roles in human complex diseases. iMeta. 2022;1(2):e14.

98. Watt E, Gemmell MR, Berry S, Glaire M, Farquharson F, Louis P, et al. Extending colonic mucosal microbiome analysis-assessment of colonic lavage as a proxy for endoscopic colonic biopsies. Microbiome. 2016;4(1):61.

99. Vaga S, Lee S, Ji B, Andreasson A, Talley NJ, Agréus L, et al. Compositional and functional differences of the mucosal microbiota along the intestine of healthy individuals. Sci Rep. 2020;10(1):14977.

100. Yan W, Sun C, Zheng J, Wen C, Ji C, Zhang D, et al. Efficacy of fecal sampling as a gut proxy in the study of chicken gut microbiota. Front Microbiol. 2019;10:2126.

101. van der Wijst MGP, Brugge H, de Vries DH, Deelen P, Swertz MA, Franke L. Single-cell RNA sequencing identifies cell type-specific cis-eQTLs and co-expression QTLs. Nat Genet. 2018;50(4):493–7.

102. Jagadeesh KA, Dey KK, Montoro DT, Mohan R, Gazal S, Engreitz JM, et al. Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. Nat Genet. 2022;54(10):1479–92.

103. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. PLoS Genet. 2021;17(9):e1009440.

Publisher's Note