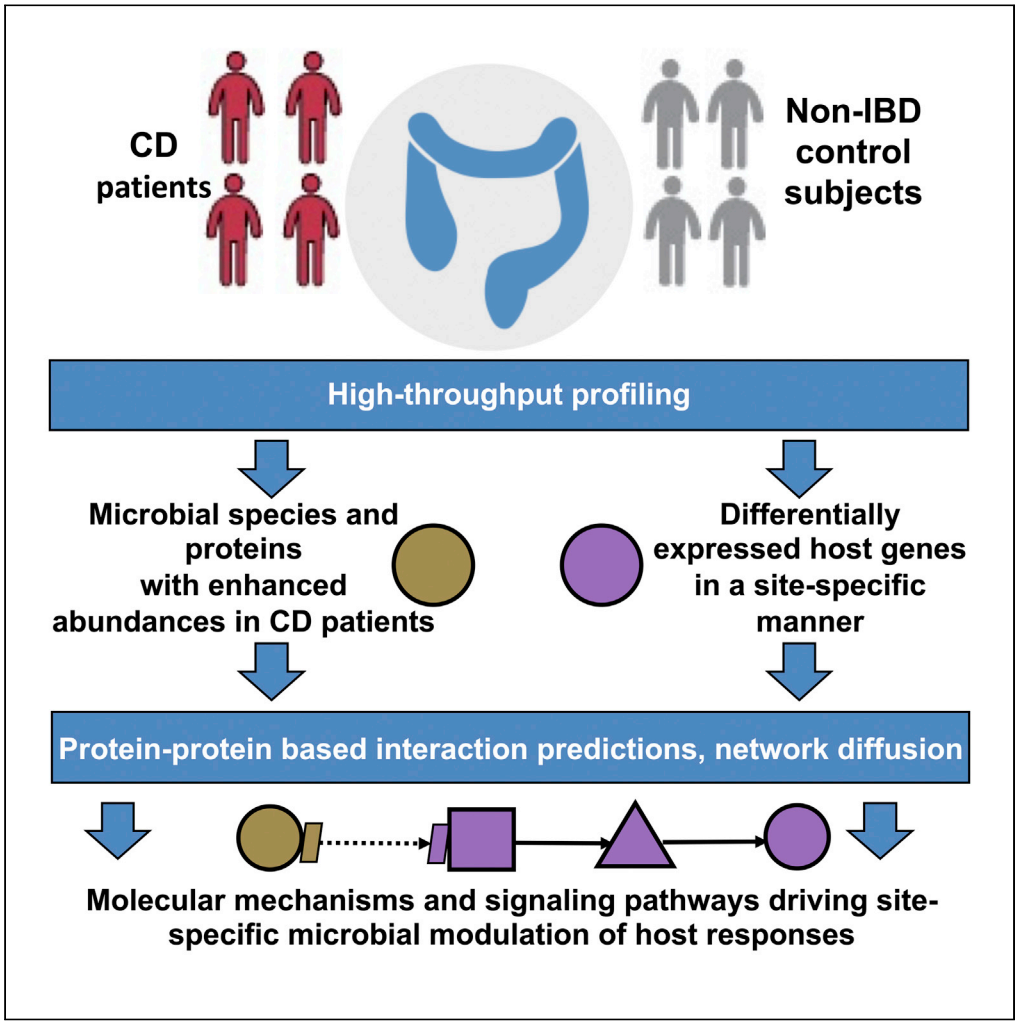
iScience

CellPress
OPEN ACCESS

Article

Integrated analysis of microbe-host interactions in Crohn's disease reveals potential mechanisms of microbial proteins on host gene expression



Padhmanand Sudhakar, Tahila Andrighetti, Sare Verstockt, ..., João Sabino, Bram Verstockt, Severine Vermeire

padhmanand.sudhakar@kuleuven.be

Highlights

Integration of community-wide microbial profiling with host transcriptomics in CD

Structural feature-based protein-protein interactions reveal microbe-host cross-talk

Site-specific microbiota mediated host signaling networks reveal key hubs and pathways

Microbiota mediated signaling networks are pathologically and therapeutically relevant

Sudhakar et al., iScience 25, 103963
May 20, 2022
https://doi.org/10.1016/j.isci.2022.103963

iScience

CellPress
OPEN ACCESS

Article

Integrated analysis of microbe-host interactions in Crohn's disease reveals potential mechanisms of microbial proteins on host gene expression

Padhmanand Sudhakar,[1,4,*] Tahila Andrighetti,[2] Sare Verstockt,[1] Clara Caenepeel,[1,3] Marc Ferrante,[1,3] João Sabino,[1,3] Bram Verstockt,[1,3] and Severine Vermeire[1,3]

SUMMARY

Inflammatory responses of the intestinal epithelial barrier in patients with Crohn's disease (CD), a chronic inflammatory bowel disease (IBD), are associated with gut microbial alterations. At a community level, there is scarce mechanistic evidence on the effects of gut microbial alterations on host mucosal barrier responses. We used a computational microbe-host interaction prediction framework based on network diffusion and systems biology to integrate publicly available paired gut microbial and intestinal gene expression datasets. The ileal signaling network potentially modulated by the microbiota was enriched with immune-related pathways such as those associated with IL-4, IL-2, IL-13, NFkB, and toll-like receptors. We identified bacterial proteins eliciting post-translational modifications on host receptors, resulting in the de-repression of pro-inflammatory cytokines via critical hub proteins such as NFkB. The signaling networks were over-represented with CD associated genes and CD drug targets. Using datasets generated from our validation cohorts, we confirmed some of the results.

INTRODUCTION

Crohn's disease (CD) is a sub-type of inflammatory bowel disease (IBD), characterized by intestinal microbial dysbiosis (DeCruz et al., 2015; Lloyd-Price et al., 2019; Pascal et al., 2017; Schaubeck et al., 2016). Although a complex set of factors including host genetics, lifestyle, diet, and exposure to antibiotics contribute to the development of CD (Ananthakrishnan et al., 2018), the microbiome-host axis is considered as a fulcrum influencing the manifestation of CD symptoms (Ananthakrishnan et al., 2017; Erickson et al., 2012; Gevers et al., 2014; Halfvarson et al., 2017; Jostins et al., 2012; Pérez-Brocal et al., 2015). This is reflected by alterations in the gut microbial compositions of IBD patients at various taxonomic levels, including enhanced abundances of particular species and epithelial barrier responses (Erickson et al., 2012; Lloyd-Price et al., 2019). Some of these responses result in hyperinflammation (Craven et al., 2012; Shaw et al., 2016) and degradation of the mucus layer which makes the epithelial barrier more vulnerable (Johansson et al., 2014; Roberton and Corfield, 1999; Wu et al., 2011, 2014; Yu, 2015; Yu et al., 2014) to microbial components. These observations are also supported by changes in the expression of genes and/or proteins involved in these processes in CD patients exhibiting microbial dysbiosis (Lloyd-Price et al., 2019).

Although several lines of evidence point to the importance of microbiome-host interactions in CD, the underlying molecular mechanisms that mediate the cross-talk between the microbiome and the host are largely unknown. This could be due to (1) the lack of comprehensive datasets with coherent/parallel sampling and high resolution profiling of the gut microbiome and intestinal gene expression from the same samples/subjects; (2) practical difficulties associated with the retrieval of enough biomaterial from luminal samples/biopsies to be able to measure multiple read-outs; (3) the inability of *in vitro* systems such as organoid cultures, cell-lines, and microfluidic systems to truly reproduce in-vivo conditions and the ensuing microbiome-host interactions; and (4) high costs and practical limitations associated with high-throughput technologies to perform live profiling of microbiome-host interactions.

Computational approaches provide an alternative option to investigate and explore the role of microbiome-host interactions (Guven-Maiorov et al., 2017). Various methods have been applied in the last decade to infer and predict the interactions between microbial and host proteins (Cui et al., 2016;

[1]KU Leuven Department of Chronic Diseases, Metabolism and Ageing, Translational Research Center for Gastrointestinal Disorders (TARGID), IBD group, ON I Herestraat 49 - box 701, 3000 Leuven, Belgium

[2]Institute of Biosciences, São Paulo University (UNESP), Botucatu 18618-689, SP, Brazil

[3]Department of Gastroenterology and Hepatology, University Hospitals Leuven, KU Leuven, Leuven, Belgium

[4]Lead contact

*Correspondence: padhmanand.sudhakar@kuleuven.be

https://doi.org/10.1016/j.isci.2022.103963

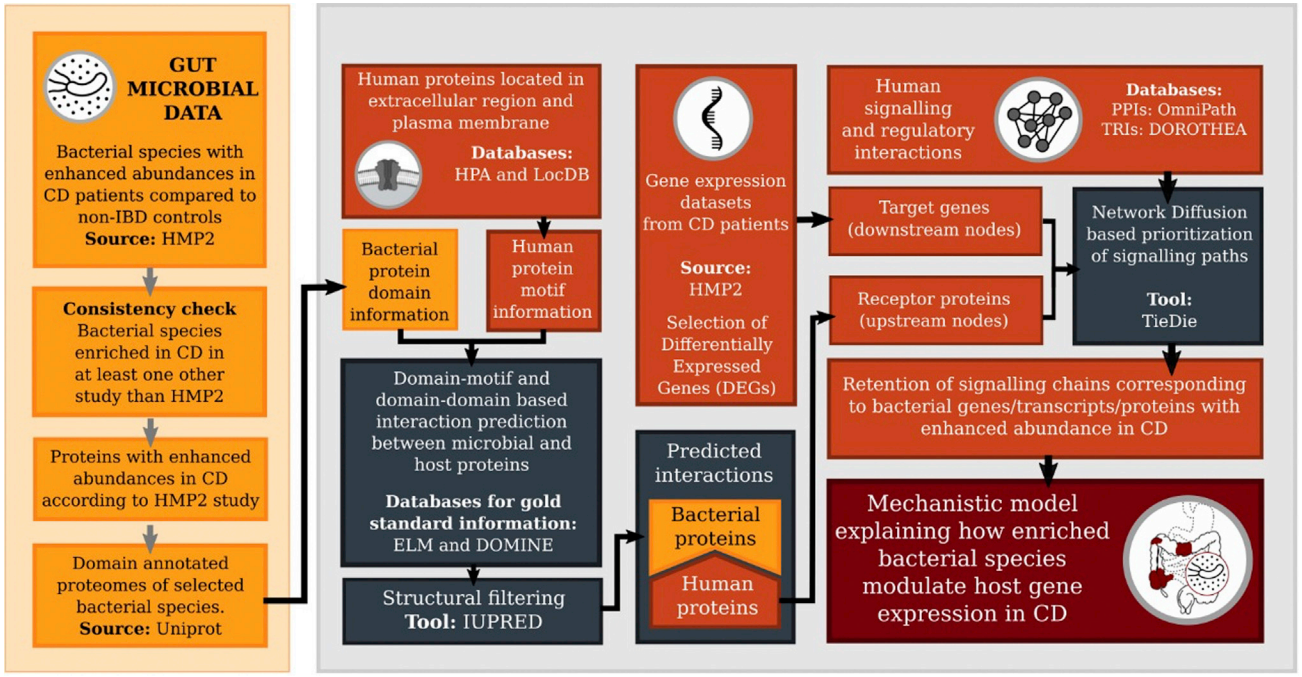


Figure 1. Graphical description of the workflow used to infer the plausible effect of the altered CD gut microbial proteins on host intestinal gene expression

Krishnadev and Srinivasan, 2011; Kshirsagar et al., 2012; Nourani et al., 2015) (Huang and Brumell, 2014; Schweppe et al., 2015; Sudhakar et al., 2019) (Bose et al., 2019). However, most of these tools are either confined solely to the upstream interactions between microbial and host proteins or document only experimentally verified protein-protein interactions corresponding to pathogens (Durmuş Tekir et al., 2013; Kumar and Nanduri, 2010; Li et al., 2015; Vialas and Gil, 2015). Other existing tools in the computational domain of microbiome-host interactions are limited to the analysis of metabolic networks (Agren et al., 2013; Arkin et al., 2018; Heirendt et al., 2019; Levy et al., 2015). Nevertheless, a common trait in most of the existing microbiome-host interaction analysis tools is that they do not consider the cumulative effects of inter-kingdom protein-protein interactions on downstream host processes. Recent tools such as MicrobioLink (Andrighetti et al., 2020), a computational pipeline investigating the functional effects of microbiome-host interactions by linking microbial-host protein interactions to the expression of host proteins using network propagation principles, enable the inference of inter-kingdom cross-talks.

Motivated by the availability of synchronous high-resolution datasets profiling the gut microbial alterations and gene expression measurements in CD patients via the IBDMDB/HMP2 project (Lloyd-Price et al., 2019), we applied a modified workflow (Figure 1) of the MicrobioLink pipeline to the IBDMDB/HMP2 dataset to infer the cumulative effects of an altered CD gut microbiome on intestinal gene expression in CD patients. Topological and functional analysis of the resulting networks aided in the identification of the prominent hubs and signaling pathways mediating the effect of the altered microbiome on differentially expressed genes in the ileum and rectum of CD patients. Using an independent cohort composed of CD patients, we validated some of our findings. We also investigated the disease relevance and druggability of the hubs as well as the gene expression signatures modulated by the altered CD microbiome.

RESULTS

Bacterial proteins altered in CD can potentially bind to human proteins

By combining information about enhanced abundances and transcriptional activity of bacterial species in CD as reported in the HMP2 IBDMDB dataset with other independent studies reporting the same trend, we identified eight bacterial species (Figure 2A, Table S1) relevant to CD. This approach based on multiple lines of evidence (enhanced abundance, transcriptional activity from one same study and additional confirmation of enhanced abundance from another study) imparts greater reliability in making inferences from

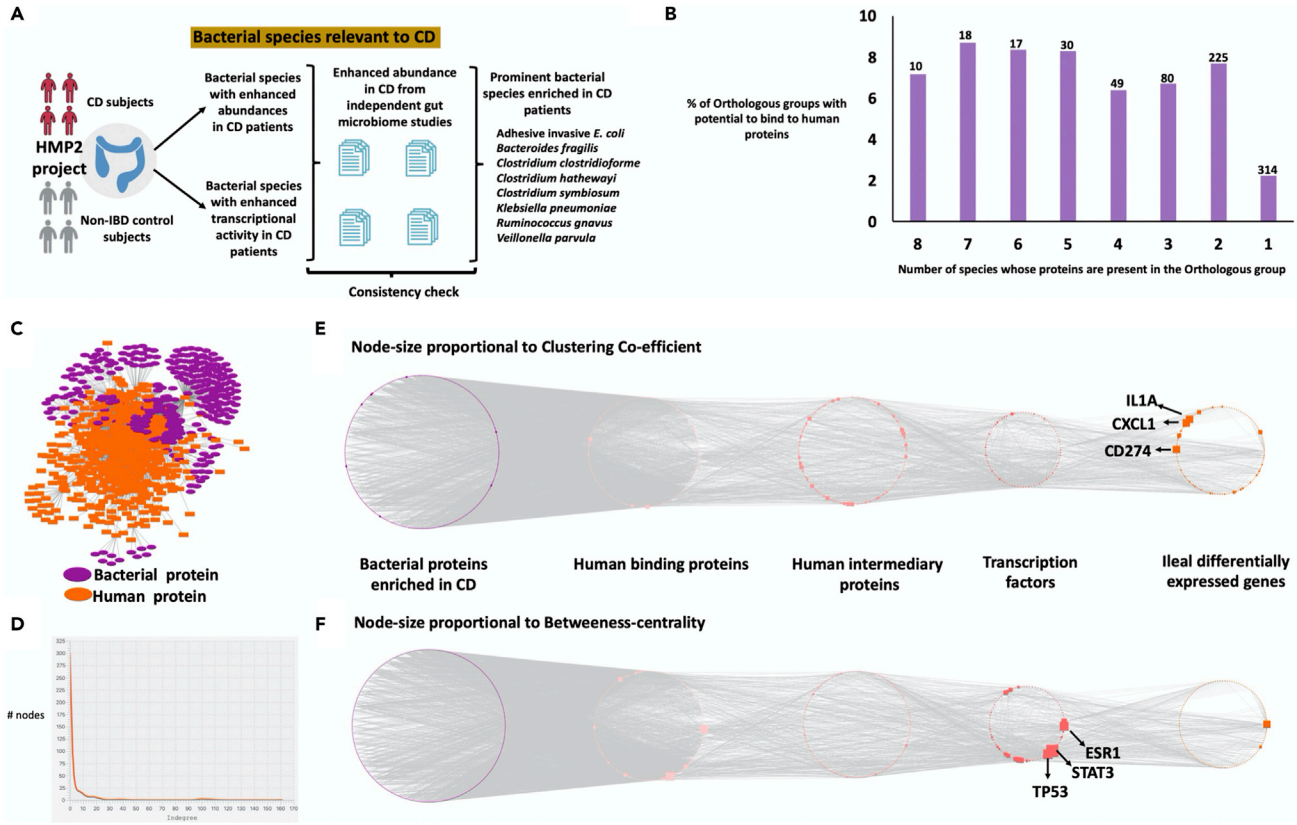


Figure 2. Overview of the stepwise results leading to the inference of the site-specific signaling networks

(A) Schematic displaying the procedure based on multiple evidences to identify bacterial species relevant to (based on enhanced abundance and transcriptional activity) CD.

(B) Figure displaying the relationship between the fraction of orthologous groups of proteins (from the 8 CD relevant species) which are capable of binding to human proteins and the number of species with protein membership in the orthologous groups.

(C) Hairball representation of the signaling network capturing the potential effect of the microbial proteins on ileal gene expression in CD patients.

(D) In-degree distribution of the ileal signaling network.

(E and F) The segregation of the ileal signaling network into sequential layers based on their position in the signaling chains mediating the effect of the bacterial proteins on ileal gene expression. The size of the nodes in E denotes the clustering co-efficient, whereas in F it stands for the betweenness centrality. The top three ranked nodes according to the clustering co-efficient are indicated in E while the top three transcription factors (TFs) ranked by the betweenness centrality are shown in (F).

noisy microbiome datasets (Figure 2A). The eight identified species include Adhesive invasive *Escherichia coli* (AIEC), *Bacteroides fragilis, Clostridium clostridioforme, Clostridium hathewayi, Clostridium symbiosum, Klebsiella pneumoniae, Ruminococcus gnavus*, and *Veillonella parvula* (Figure 2A, Table S1). By restricting the microbe-host protein-protein interaction prediction (see STAR Methods section for more details) to the homologs (in the eight CD relevant species) of enhanced microbial genes and transcripts (Table S2) identified in CD patients in the HMP2 IBDMDB cohort, we inferred 739 orthologous groups with the potential ability to bind to human proteins (Figure 2B, Table S3). Limiting the microbe-host protein-protein interaction prediction to microbial genes and transcripts significantly enhanced in CD proteins further ensures the specificity of the microbial proteins with respect to their role in CD pathogenesis. The 739 orthologous groups thus identified make up about 3.7% of the total number of orthologous groups identified by clustering all the proteins from the eight bacterial species (Figure 2B). About 73% (539) of the 739 orthologous groups had protein members from either one or two species only which could indicate that a large proportion of protein-clusters are specific to particular species. In contrast, only ten orthologous groups (1.3%) contained proteins from all the eight species, suggesting that generic effects at the level of microbe-host PPIs are minimal compared to species specific effects (Figure 2B). Functional analysis (hypergeometric test; FDR $\leq$ 0.05) of the 739 orthologous groups with potential binding capabilities identified several metabolic functions in addition to processes involved in post-translational modifications.

Network diffusion highlights the potential modulation of host gene expression in CD by bacterial proteins

Even though we identified bacterial proteins capable of binding to human proteins, it does not necessarily imply that these interactions have an influence on disease development or progression. To fill this gap, we harnessed the power of network diffusion to investigate if and how the inferred bacterial-host PPIs can have downstream effects as measured by gene expression from the host. Using this approach, our analysis revealed some of the potential mechanisms and key regulators that could mediate the effect on host gene expression of the altered bacterial proteins in CD. The signaling network (Figure 2C), resembling a scale-free network of 901 nodes and 1600 interactions (Figure 2D), captures the potential effect of the altered bacterial proteins on ileal gene expression in CD patients. The prominent components as inferred by network topological metrics such as betweenness centrality (the extent to which a node is a bridge between two other nodes) and clustering co-efficient (the extent to which the neighbors of a node are connected to each other) (Figure 2F, Table S4) point to proteins known to play important functional roles in intestinal inflammation, immune interactions/signaling and often with expression/activity modulations in CD patients (Tables 1 and S4). Top ranked nodes by betweenness centrality include critical transcription factors *STAT3* (UniProt: P40763), *TP53* (UniProt: P04637), *ESR1* (UniProt: P03372), *STAT1* (UniProt: P42224), *SPI1* (UniProt: P17947), *SMAD3* (UniProt: P84022), and *MYC* (UniProt: P01106) (**Table 1**) in addition to signaling proteins represented by *SOCS3* (UniProt: O14543), *SRC* (UniProt: P12931), *NCOA1* (UniProt: Q15788), *CTNNB1* (UniProt: P35222), and *PRKACA* (UniProt: P17612) among others (Tables 1 and 4).

The human signaling network (i.e., excluding the bacterial proteins) capturing the effect of altered bacterial proteins on ileal gene expression in CD patients was over-represented with Reactome pathways such as signaling mediated by interleukins (IL-4 (UniProt: P24394), IL-2 (UniProt: P60568), and IL-13 (UniProt: P35225)), receptor tyrosine-kinases, multiple toll-like receptors including *TLR4* (UniProt: O00206), *MyD88* (UniProt: Q99836), *NFkB* (UniProt: P19838) (*TRAF6* mediated induction of *NFkB* upon *TLR7/8* or *9* activation), in addition to FCERI mediated MAPK activation (Figures 3A, 3B and Table S5). Meanwhile, functional over-representation analysis confined to genes differentially expressed (all upregulated - Figure 3C) in the ileum concurrently point to interleukin (*IL-4,10,13*) signaling and immune system functions among other generic functions such as gene expression and transcriptional control (Figures 3C and S1). Using an independent cohort of CD patients, we also validated the microbiome-induced gene expression in the ileum of CD patients in our test cohort. We observed that 78% (92/117) of the DEGs in our test cohort were also differentially expressed in the validation cohort (R2 = 0.907) (Figure 3D, Table S6).

The ileal signaling network with both bacterial and host proteins could be broken down into three distinct modules based on the overall connectivity with other proteins in the network (Figure 3E) with common and specific functions (Figure 3F, Table S7).While modules 1 and 2 shared several over-represented pathways such as *TRIF*-mediated *TLR4* signaling, *IL-4/IL-13* signaling among many others, several module specific pathways could also be identified. For example, module 1 was specific to the enrichment of interferon and *PI3K/AKT* signaling while *FCERI* (UniProt: P12319) and *TGF-beta* receptor signaling were confined to module 2. Module 3 was, however, peculiarly distinct in terms of its composition by harboring genes corresponding to exclusively enriched pathways such as response to infection by parasites, *FCGR-3A* mediated phagocytosis, cell-cell communication, *BRAF/RAF1* UniProt: P15056) signaling, *ERBB2* (UniProt: P04626) signaling among others (Figure 3F, Table S7).

As potential drivers of the ileal gene expression changes described above, several bacterial proteins (which could be clustered into groups based on sequence homology) could be identified to have an influence as shown in Figure 3G. These include proteins with acetyltransferase, chaperone, peptidase, peptidyl-prolyl *cis*-trans isomerase and proteolytic functions among others. For example, the orthologous group OMA01692, representative of a peptidase S8 domain containing group of proteins present in 4 (*C. clostridioforme, C. hathewayi, C. symbiosum, R. gnavus*) of the 8 CD relevant bacterial species, potentially influences the expression of 43% of the differentially expressed genes in the ileum of CD patients.

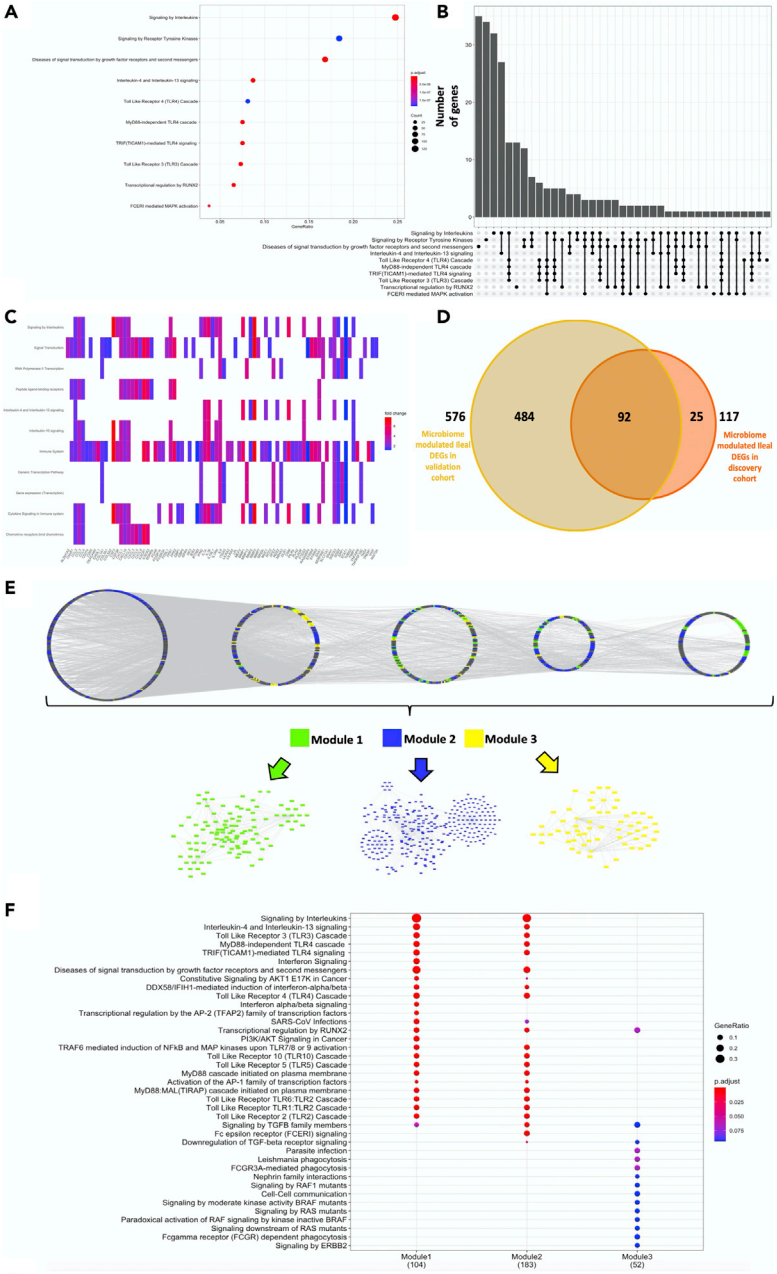
Regulatory control which could be exerted by bacterial influences can potentially induce site-specific inflammatory responses at the level of gene expression in CD

We also observed that the bacterial species identified as being relevant in CD tend to modulate ileal gene expression through prominent TFs such as *SPI1*, *STAT1* and NFKB1 (UniProt: P19838) as identified by the network topology analysis (Figure 4A, Table S8). Based on our results, we inferred that the activity of these

Table 1. The functions (in intestinal inflammation and IBD/CD) of the top 10 proteins (as inferred by the metric of betweenness centrality) in the ileal signaling network

Node	Protein name	UniProt biological process keywords	Role in intestinal inflammation	Known role in CD/IBD
STAT3	Signal transducer and activator of transcription 3	Host-virus interaction, Transcription, Transcription regulation	STAT3 phosphorylation protects against intestinal inflammation(Li et al., 2017)[a], STAT3 modulates intestinal inflammatory response(Kim et al., 2016)[a], Inactivation of IL-6/STAT3 cascade contributes to the attenuation of chronic intestinal inflammation(Atreya et al., 2000)	STAT3 Activation confined to actively inflamed colons in CD patients(Musso et al., 2005), increased total STAT3 in CD(Mudter et al., 2005), increased pSTAT3 correlated with histological degree of inflammation(Mudter et al., 2005)
TP53	Cellular tumor antigen p53	Apoptosis, Biological rhythms, Cell cycle, Host-virus interaction, Necrosis, Transcription, Transcription regulation	Promotes intestinal inflammation(Cooks et al., 2013; Spehlmann et al., 2013)[a]	Most frequent mutation in IBD-associated dysplastic lesions and in cancers(Alpert et al., 2019; Wanders et al., 2020), Nested or diffuse TP53 in noncancer IBD(Laurent et al., 2011)
ESR1	Estrogen receptor	Transcription, Transcription regulation	Positive regulator of intestinal inflammation(Goodman et al., 2020)[a]	Modulates mucosal inflammation in IBD(Cook et al., 2014)[a]
SOCS3	Suppressor of cytokine signaling 3	Growth regulation, Ubl conjugation pathway	Positive regulator of intestinal inflammation(Koay et al., 2014; Suzuki et al., 2001)	Modulator of intestinal inflammation in CD (25,997,679), Elevated in CD subgroups with active inflammation at both ileum and colon(Savić Mlakar et al., 2018)
SRC	Proto-oncogene tyrosine-protein kinase Src	Cell adhesion, Cell cycle, Host-virus interaction, Immunity	Involved in linking inflammation to intestinal epithelial regeneration(Taniguchi et al., 2015)[a]	–
STAT1	Signal transducer and activator of transcription 1-alpha/beta	Antiviral defense, Host-virus interaction, Transcription, Transcription regulation	Tumor suppressor molecule in inflammation-associated carcinogenesis(Leon-Cabrera et al., 2018)	Enhanced STAT1 expression in lamina-propria T-cells and colonic biopsies of CD patients(Taniguchi et al., 2015; Wu et al., 2007)
SPI1	Transcription factor PU.1	Transcription, Transcription regulation	Modulates chronic intestinal inflammation(Kekuda et al., 2008)[a] and shigellosis(Ta et al., 2017)[a]	Upregulated in patients with Crohn'l's ileitis(Fagnart et al., 1985)
SMAD3	Mothers against decapentaplegic homolog 3	Host-virus interaction, Transcription, Transcription regulation	–	SMAD3 phosphorylation reduced in CD(Monteleone et al., 2001)[a]
MYC	Myc proto-oncogene protein	Transcription, Transcription regulation	MYC dependent deregulation of Wnt signaling promotes carcinogenesis(Shah et al., 2015)	High frequency of MYC amplification in IBD associated intestinal adenocarcinomas(Hartman et al., 2018), Reduced expression in IECs of IBD patients(Alexander et al., 1996)
NCOA1	Nuclear receptor coactivator 1	Transcription, Transcription regulation	–	–

[a]Evidence from mice experiments.



G

Ortholog_id	Representative function	% Ileal DEGs reached	# Bacterial species*	EC	BF	CC	CH	CS	KP	RG	VP
OMA01963	Predicted metal-binding enzyme	7.7	8								
OMA02114	Peptidyl-prolyl cis-trans isomerase	28.6	6								
OMA02791	Acetyltransferase	26	5								
OMA01692	Peptidase_S8 domain-containing protein	43	4								
OMA00345	DegP-like periplasmic serine endoprotease	43	3								
OMA01126	Ferrochelatase	38.1	3								
OMA01566	Peptidyl-prolyl cis-trans isomerase	5.1	3								
OMA02172	Peptidyl-prolyl cis-trans isomerase	28.6	3								
OMA00962	Universal stress protein	42.5	2								
OMA01035	DegP-like periplasmic serine endoprotease	42.8	2								
OMA02061	Serine protease	26	2								
OMA02099	ATP-dependent protease subunit	38.5	2								
OMA02361	Universal stress protein	42.5	2								
OMA02671	Co-chaperone protein	41.7	2								
OMA02733	Peptidyl-prolyl cis-trans isomerase	5.1	2								
OMA03462	Universal stress protein	42.5	2								

% Total Ileal DEGs
5 to 12
13 to 20
21 to 28
29 to 36
37 to 43

Figure 3. Functional analysis of the microbiota-modulated ileal signaling network

(A) Top 10 Reactome pathways enriched (adj.p-value ≤ 0.1) in the ileal signaling network (B) and overlap between them.

(C) Expression profiles of the genes belonging to the top 10 Reactome pathways enriched (adj.p-value ≤ 0.1) among the differentially expressed genes in the ileal signaling network.

(D) Overlap between the ileal DE-Gs potentially modulated by the altered microbiota in CD patients and DE-Gs measured in an independent cohort of CD patients.

(E) Modules (tightly connected networks) detected by MCODE within the ileal signaling network.

(F) Over-represented Reactome signaling pathways (adj.p-value ≤ 0.1) in the network modules inferred from the ileal signaling network.

(G) Bacterial proteins with post-translational modification functions and inferred to influence the differential expression of genes in the ileum. EC: Adhesive invasive *E. coli* (AIEC); BF: *Bacteroides fragilis*; CC: *Clostridium clostridioforme*; CH: *Clostridium hathewayi*; CS: *Clostridium symbiosum*; KP: *Klebsiella pneumoniae*; RG: *Ruminococcus gnavus*; VP: *Veillonella parvula*. * - Number of bacterial species in which a protein member from the ortholog group was detected. Singleton proteins or ortholog groups with just membership from a single species have not been shown in this figure.

TFs could be potentially influenced by multiple species with most of the TFs among the top 20 betweenness centrality ranked nodes capable of being modulated by at least seven of the eight CD relevant species. *SPI1*, *STAT1* and *NFKB1* individually regulate the expression of 23%, 20 and 17% of the DE-Gs and cumulatively around 52% of the non-redundant DE-Gs in the ileal signaling network (Figure S3). More than two-thirds (71%) of the DE-Gs modulated by the top TFs are also involved in immune responses and inflammation. *NFKB1*, for example, is indirectly modulated upstream by the S8 domain containing peptidase cleaving *MAP3K7* (UniProt: O43318) by binding to the PCSK cleavage motif CLV_PCSK_PC1ET2_1 (Figures 4B and 4C). Inactivation of MAP3K7 leads to the activation of NFKB1 which subsequently results in the upregulation of several genes with functions related to inflammation (*IL6* (UniProt: P05231), *IFNG* (UniProt: P01579), etc.), recruitment and adhesion of immune cells (*CCL2* (UniProt: P13500), *CXCL10* (UniProt: P02778), etc.), remodeling of the extracellular matrix (*MMP13* (UniProt: P45452), *COL3A1* (UniProt: P02461), etc.) and enzymes involved in metabolizing inflammatory pathway precursors (PTGS2). This is one of several examples wherein the influence of bacterial proteins results in the upregulation of genes involved in pathogenesis of CD. With the exception of four genes (*CCL2*, *COL1A1* (UniProt: P02452), *COL3A1*, *MMP2* (UniProt: P08253)), all the other genes under the transcriptional control of three TFs including NFKB1 (which mediate the modulation by the S8 domain containing peptidase) are differentially expressed in the validation cohort (R2 = 0.803) (Figure 4D). When all the ileal DE-Gs under the control of the top TFs among the top 20 betweenness centrality ranked nodes were considered together, we observed a consistent trend between the test and validation cohorts (R2 = 0.908).

We also compared the regulatory footprints across the ileal and rectal signaling networks (Figures S4A and S4B) and found several commonalities and differences (Figure 4E, Table S8) at the level of regulation of DE-Gs. Of the 87 DE-Gs regulated by a TF from among the set of top 20 nodes (based on betweenness centrality) of the ileal and rectal signaling networks, 54% (47/87) were common to both the ileum and rectum. In other words, ileum and rectum not only harbor common gene signatures in CD but could also be controlled by a set of shared regulators that could be driving the synchronous gene expression. With an analysis of regulators with at least 10 target DE-Gs in either ileum or rectum, we observed different levels of specificity (ranging from 35 to 61%) not only quantitatively but also qualitatively. For example, the DEG targets (COL1A1, MMP13, MMP2, PLAUR (UniProt: Q03405), CCL2, SOD2 (UniProt: P04179), TFPI2 (UniProt: P48307), TIMP1 (UniProt: P01033)) of *NFKB1* exclusive to ileum are associated with extracellular matrix remodeling and organization. Conversely, the targets of NFKB1 common to both ileum and rectum are predominantly associated with immune and inflammatory functions. Furthermore, we also identified pathways and biological processes among the genes in the potential microbiota-mediated signaling networks in a site-specific manner (Figure 4F, Table S5).While processes associated with anti-bacterial defense, cellular fate and cellular adhesion among others were over-represented in the ileal signaling network, the rectal network was characterized by the enrichment of metabolic processes. At the level of Reactome signaling pathways, ileum-specific ones include those related to collagen formation, assembly and degradation as well as platelet adhesion to exposed collagen structures.

Disease relevance and druggability of the potential microbiota-modulated signaling networks

To reveal the relevance of the inferred signaling networks modulated by the bacterial proteins, we checked if the proteins in the networks are associated with CD as reported in the OpenTargets database

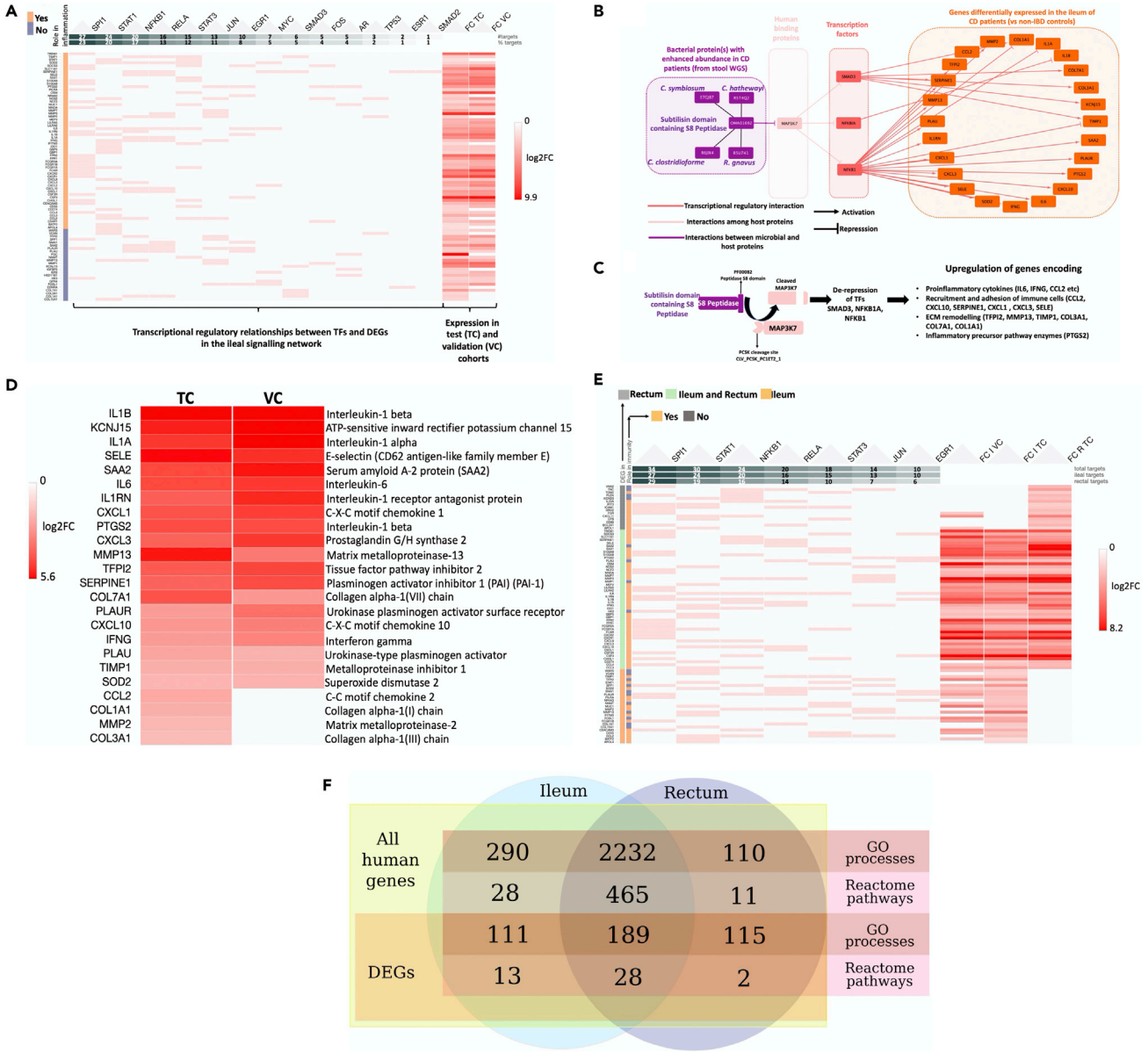


Figure 4. Examples of regulatory and mechanistic interactions mediated by microbial proteins followed by the comparison of the test and validation cohorts

(A) Heatmap representation of the transcriptional factors (among the top 20 nodes in the ileal signaling network ranked by the betweenness centrality) regulating the DE-Gs in the ileal signaling network. The number of targets for each of the TFs and the involvement of the DE-Gs in immune responses are depicted. The FC TC and FC VC columns stand for the expression level (in log2 scale) of the DE-Gs from the test and validation cohorts respectively.

(B and C) Inferred molecular mode of action of the subtilisin-domain containing type S8 peptidase (encoded by four different species with enhanced abundances in CD) on ileal gene expression in CD patients. The potential cleavage of *MAP3K7* by the peptidase in turn enhances the activity of transcription factors (*SMAD3*, *NFKB1*, *NFKBIA*) which subsequently results in the upregulation of genes which are involved in inflammation, recruitment and adhesion of immune cells, remodeling of extracellular matrix and encoding enzymes involved in the production of inflammatory precursors.

(D) Comparison between the test (TC) and validation cohorts (VC) of the log2 normalized expression fold changes ($1.5 \leq \log2FC \leq -1.5$, FDR ≤ 0.05) of the transcriptional regulatory targets of *SMAD3*, *NFKB1*, *NFKBIA*which are modulated by the subtilisin-domain containing type S8 peptidase.

(E) Heatmap representation of the transcriptional factors (among the top 20 nodes in the ileal and rectal signaling networks ranked by the betweenness centrality) regulating the DE-Gs in the ileal and rectal signaling networks. TFs with less than 10 targets each among the ileal and rectal DE-Gs are not shown. The number of targets specific to ileum/rectum and common to both ileum and rectum for each of the TFs is depicted. The FC I TC and FC I VC columns stand for the expression level (in log2 scale) of the ileal DE-Gs from the test and validation cohorts respectively. FC R TC stands for the expression level of the DE-Gs in the rectal signaling network.

(F) Summary of the over-represented gene ontological biological processes and Reactome pathways in the microbiome-modulated ileal and rectal signaling network.

Table 2. Ileal and rectal signaling network proteins which are also targeted by drugs in CD

Protein symbol	Protein name	# active/completed clinical trials	Drug(s)
BMX	Cytoplasmic tyrosine-protein kinase BMX	1	PF-06651600
BTK	Tyrosine-protein kinase BTK	1	PF-06651600
CD80[b]	T-lymphocyte activation antigen CD80	–	–
CSF2RA	Granulocyte-macrophage colony-stimulating factor receptor subunit alpha	2	Sargramostim
CSF3R	Granulocyte colony-stimulating factor receptor	3	Filgrastim
CXCL10	C-X-C motif chemokine 10	1	Eldelumab
ESR2	Estrogen receptor beta	1	Prinaberel
GHR	Growth hormone receptor	1	Somatropin
IL2RB	Interleukin-2 receptor subunit beta	1	Aldesleukin
IL6	Interleukin-6	2	PF-04236921
IL23A[b]	Interleukin-23 subunit alpha	12	Brazikumab, Risankizumab, Guselkumab
ITGA4	Integrin alpha-4	30	Vedolizumab, Natalizumab, Abrilumab
JAK1	Tyrosine-protein kinase JAK1	15	Upadacitinib, Tofacitinib, Filgotinib
JAK2	Tyrosine-protein kinase JAK2	9	Upadacitinib, Tofacitinib
JAK3	Tyrosine-protein kinase JAK3	10	Upadacitinib, Tofacitinib, PF-06651600
MMP1	Interstitial collagenase	2	Doxycycline
MMP7	Matrilysin	2	Doxycycline
MMP13[a]	Collagenase 3	2	Doxycycline
NR3C1	Glucocorticoid receptor	13	Budesonide, Methylprednisolone, Prednisolone, Prednisone
PPARG	Peroxisome proliferator-activated receptor gamma	7	Mesalamine, Pioglitazone
PTGS2	Prostaglandin G/H synthase 2	7	Mesalamine
RBX1	E3 ubiquitin-protein ligase	4	Thalidomide, Lenalidomide
TYK2	Non-receptor tyrosine-protein kinase	9	Upadacitinib, Tofacitinib
VDR	Vitamin D receptor	10	Ergocalciferol, Cholecalciferol, Calcitriol

[a]Protein present only in ileal signaling network.
[b]Protein present only in rectal signaling network.

(Carvalho-Silva et al., 2019). About 22.76% (125/549; enrichment p-value: 1.38E-15) and 23.42% (115/491; enrichment p-value: 2.24E-15) of the host proteins in the ileal and rectal signaling networks respectively were annotated as being associated with CD (Table S9) based on at least two sources of evidence (see Table S12 for the full filtered list of CD associated OpenTargets proteins). Among the OpenTargets annotated CD proteins in the ileal and rectal networks, 46/125 (37%) and 40/115 (35%) respectively are derived from evidence sources with a genetic basis although information on host genotypes was not integrated in the current study. In addition, by using drug target information from OpenTarget, some of the proteins in the ileal and rectal signaling networks are known targets of therapeutic molecules used to treat CD (Table 2). Cumulatively, 24 proteins from the ileal and rectal signaling networks are existing targets of CD drugs. These include prominent cytokines such as IL23B (specific to the rectal network), *IL6* and *CXCL10* and chemokine receptors (*CSF2RA* (UniProt: P15509), *CSF3R* (UniProt: Q99062)) (common to both networks). The therapeutically targeted proteins also cover other functional categories such as those associated with JAK-STAT signaling (*JAK1* (UniProt: P23458), *JAK2* (UniProt: Q60674), *JAK3* (UniProt: P52333), *TYK2* (UniProt: P29597)), matrix organization and remodelling (*MMP1* (UniProt: P03956), *MMP7* (UniProt: P09237), *MMP13*, *ITGA4* (UniProt: P13612)), T-cell proliferation and activation (CD80 (UniProt: P33681)) and enzymes involved in the production of inflammatory precursors (PTGS2).

In addition to the above demonstrated druggability of the potential microbiota modulated signaling network components (irrespective of status of differential gene expression), we also investigated if the

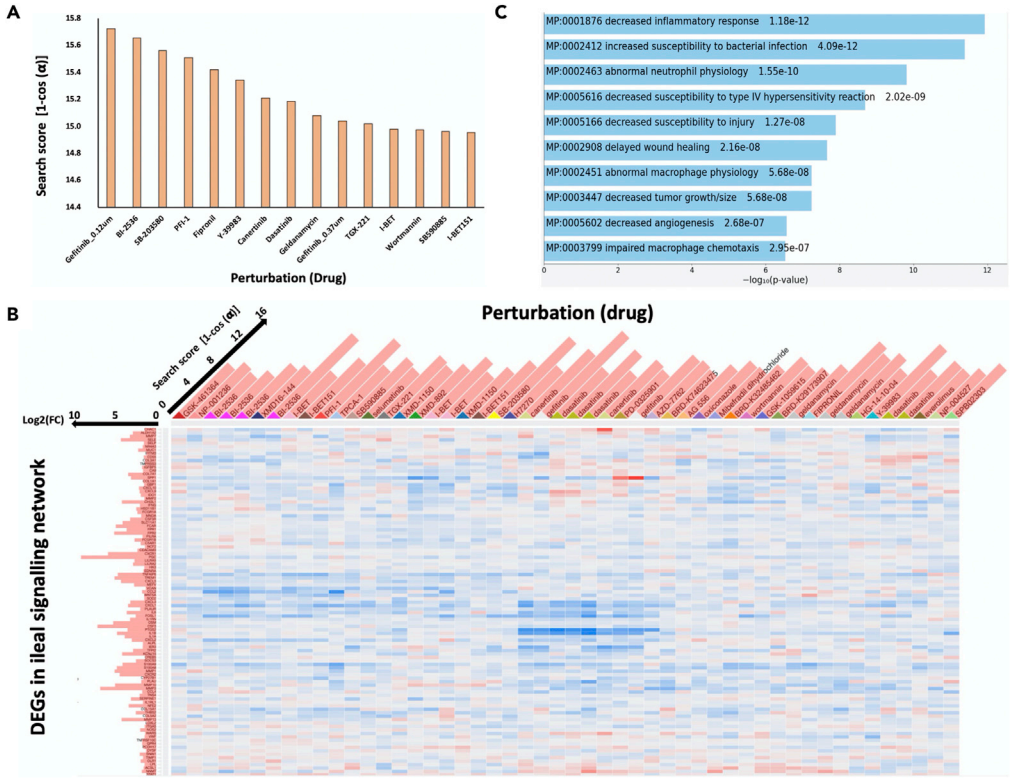


Figure 5. Expression signature based matching of drugs to targets in the ileal signaling network

(A) Search scores representing the extent to which a drug can induce/reverse the gene expression signatures of the DE-Gs in the ileal signaling network. Only the top 15 perturbation/drugs are shown.

(B) The expression profiles of the ileal network DE-Gs in response to various perturbation/drugs. The horizontal red-bars along the rows indicate the extent to which the corresponding ileal DEG is differentially expressed in CD patients. The red-label bars along the columns indicate the search score which is representative of the extent to which the perturbation/drug elicited gene expression can induce/reverse the signature profile of the DEG. The red and blue squares of the matrix indicate the upregulation and downregulation of the DEG in response to the corresponding perturbation/drug.

(C) Enriched (adj P-value $\leq$ 0.05) ontology terms (from the Mammalian Phenotype) of the ileal network DE-Gs which are also overlapping with the Wortmannin elicited gene expression signature.

DE-Gs themselves could be potentially targeted directly or indirectly by existing drugs. We inferred this by matching gene expression signatures of various drugs in the LINCS database to the gene expression signatures elicited in the ileum (Figures 5A and 5B, Table S10) and rectum (Figures S6A and S6B, Table S10) of CD patients potentially by the altered microbiota. Drugs which could induce a reversal of the expression patterns of the DE-Gs in the ileum or rectum include among others gefitinib, canertinib, dasatinib, trimpiramine, enalapril, and wortmannin. As an example, wortmannin influences the expression of genes involved in various phenotypes associated with inflammatory response, susceptibility to bacterial infection, wound healing and macrophage physiology/chemotaxis (Figure 5C).

DISCUSSION

Multiple studies have demonstrated that dysbiosis and alterations of the gut microbiome are associated with phenotypic responses including inflammation of the gut mucosal barrier in CD patients. However, at a meta-level, there is a gap in terms of the molecular mechanisms that mediate the effect of the altered gut microbiome on host responses. In this study, we used a computational pipeline (harnessing the power of microbe-host protein-protein interaction predictions, biological networks and network diffusion concepts) to integrate heterogeneous-omic datasets including whole genome shotgun sequencing of the fecal microbiome and transcriptomic read-outs from ileal and rectal biopsies of CD patients. We identified site-specific signaling networks for ileum and rectum, which may explain how the altered CD gut microbiome could modulate host gene expression.

To start with, we identified eight bacterial species (AIEC, *B. fragilis, C. clostridioforme, C. hathewayi, C. symbiosum, K. pneumoniae, R. gnavus* and *V. parvula*) associated with the gut microbiome of CD patients compared to non-IBD controls. Although strain level resolution was not possible, species level resolution is expected to provide interesting insights, particularly given that (1) the inferences are based on three different evidence types and (2) only a subset of relevant genes/proteins with increased activities in CD patients are considered for the microbe-host interaction analysis. All of the eight species are either known pathogens or support the growth and survival of pathogens. They are mainly confined to the gut although non-gut niches have also been documented for some of them (Elsayed and Zhang, 2004; Knapp et al., 2017). Conversely, pathogens such as *K. pneumoniae*which are predominantly found in non-gut niches like the respiratory tract have also been identified as being relevant to CD. This is understandable since *K. pneumoniae* is known to be a pathogen which has mastered the art of evading host immune defense mechanisms in different immune cells as well as a wide range of organ systems (Bengoechea and Sa Pessoa, 2019). However, despite the fact that depletion of commensals could have an impact on intestinal homeostasis and IBD(Koboziev et al., 2014), since the goal of the study was to infer the functional effects of the microbiome on the host, we considered, for our global analysis, only microbial species and genes/transcripts with enhanced abundances and activities respectively.

Next, we identified bacterial protein groups involved in modulating the activity of human proteins in CD. The small fraction (3.7%) of such protein groups as a proportion of all possible protein clusters could possibly be attributed to various underlying reasons such as (1) the prevalence of mechanisms along other-omic layers such as metabolism in mediating microbe-host interactions (Metwaly et al., 2020) (Franzosa et al., 2019) (2) the ability of singular bacterial proteins or a small number of them to render noticeable phenotypic changes on the host (Carroll and Maharshak, 2013; Cohavy et al., 2000; Dalwadi et al., 2001) and (3) the lack of structural annotation (especially the lack of domain annotations) given that the interaction prediction between the microbial and host proteins is heavily reliant on such annotations. Nevertheless, based on their functional annotations, we identified several relevant host-interacting bacterial proteins capable of inducing post-translational effects which play a role in the pathogenesis of many chronic diseases including IBD (Caruso et al., 2019; Gordon et al., 2020; Schweppe et al., 2015; Zhang et al., 2018, 2020a).

The analytical pipeline uncovered the corresponding site-specific (ileum/rectum) signaling networks that captured the effects of the CD relevant bacterial proteins on differentially expressed host genes. Using an independent validation cohort of CD patients carrying the pathogenic B2 enterotype, we could verify our findings as exemplified by overall consensus (78%, R2 = 0.908) between the ileal DE-Gs in the test and validation cohorts. Ranking the proteins using topologically relevant parameters such as betweenness centrality (indicative of the number of shortest paths passing through the protein) (Kincaid and Phillips, 2011) resulted in the identification of central TFs and post-translational signaling proteins. The prevalence of TFs among the BC-ranked nodes is understandable given that upstream signals modulate TF activity which thus subsequently impacts the expression of their immediate targets downstream (Karin and Smeal, 1992). Thus, TFs act as signal amplifiers by modulating gene expression and are involved in complex regulatory circuits which are required for quick and adaptive responses in biological systems (Gao et al., 2018). Some of the top ranked TFs such as *STAT3*, *STAT1*, *TP53*, *ESR1*, *STAT1*, *SPI1*, *SMAD3* and *MYC* are involved in modulating intestinal inflammation and/or known to be playing a role in CD/IBD. *STAT3* and *STAT1* for example are of interest due to the multiple mechanisms by which STATs impact IBD pathogenesis. *STAT1* is differentially expressed in ileal biopsies of CD patients and its deficiency enhances predisposition to gut inflammation (Leon-Cabrera et al., 2018). Adding to the more complex role of *STAT3* in CD is its transient/dynamic involvement independent of genetic aberrations (Ferguson et al., 2010; Willson et al., 2012).*STAT3* is involved not only in mediating inflammatory responses to infections by pathogens (Backert et al., 2014) and generation of inflammatory Th17 cells (Durant et al., 2010; Liu et al., 2008) but also in modulating the return of the epithelium to steady-states following inflammatory insults (Goldsmith et al., 2011). Synonymous with our observation of *STAT3* as a major regulatory hub is the identification as a signaling hub of *SOCS3* - a partner in crime with *STAT3* in mediating intestinal inflammation. SMAD3, another hub, is involved in the TGF-beta driven inflammatory responses in CD (Monteleone et al., 2001), mucosal immunity and T-cell activation (Yang et al., 1999). In addition to the hubs, many of over-represented signaling pathways in the microbiome-modulated ileal and rectal signaling networks were also identified in independent experimental studies. For instance, *TLR3* (UniProt: O15455)*/TLR4* associated signaling pathways which were over-represented in the ileal network, have been demonstrated to be active in CD patients (Cario and Podolsky, 2000). TLR4 is also over-expressed in ileal IECs of CD patients (Silva et al., 2008) and upregulated during intestinal inflammation (Hausmann et al., 2002). This is intriguing since

TLR pathways, especially those associated with TLR4 are generally activated by bacterial lipopolysaccharides (LPS) in disease states and are also involved in mediating host response to microbial effects (van Bergenhenegouwen et al., 2016; Cerovic et al., 2009; Kløve et al., 2020). However, our results suggest novel modes of TLR activation mediated by bacterial proteins in the context of CD.

Concomitant with observations from previous studies about the role of *NFKB1* in mediating intestinal inflammatory responses in healthy states (Ramakrishnan et al., 2019) and CD (Han et al., 2017; Schreiber et al., 1998; Segain et al., 2000; Sorrentino, 2017), our results indicated that (1) NFKB1 was among the central regulators in the ileal signaling network and (2) the expression of many inflammatory cytokines was upregulated via *NFKB1* by bacterial proteins including those with proteolytic functions as demonstrated by the S8 domain containing peptidase example. Although many stimuli such as butyrate are known to inhibit inflammatory responses via *NFKB1* (Segain et al., 2000), the exact mechanisms have remained elusive, particularly in whether and how bacterial proteins can influence inflammatory responses via *NFKB1*. Under homeostatic conditions, NFKB1 under the influence of active upstream repressors such as MAP3K7 tends to have reduced activity and thereby results in the suppression of pro-inflammatory molecules. However, bacterial proteins such as the S8 domain containing peptidase present in four of the eight CD relevant species (*C. clostridioforme, C. hathewayi, C. symbiosum, R. gnavus*) potentially inactivates *MAP3K7* by binding to the PCSK cleavage motif CLV_PCSK_PC1ET2_1 and thereby inactivating *MAP3K7*. This results in downstream effects culminating in the de-repression of TFs including *NFKB1*which is then able to activate the expression of pro-inflammatory cytokines such as *IL6*, *IFNG*, *CCL2* among others. Interestingly, the S8 domain containing peptidase in absent in all but one proteome (Faecalibacterium sp. CAG:1138; protein uniprot id - R5FTW4) within Faecalibacterium sp and absent in all of the proteomes in Akkarmensia sp. Thus, bacterial proteins with enhanced abundances and activity in CD and mostly absent in commensals such as Faecalibacterium sp. and Akkarmensia sp which are depleted in CD, can "switch-off" mechanisms that under homeostatic conditions would have silenced inflammatory cytokines. That being said, the above described mechanism of action could only be one among many other microbiome-mediated mechanisms modulated by other bacterial proteins and several other players such as small RNAs and metabolites (Valter et al., 2020; Wlodarska et al., 2015).

By comparing the ileal and rectal signaling networks mediating the possible microbial influences on ileal and rectal gene expression, we identified site specific commonalities and differences in the gene expression as well as the underlying regulatory programs. At the level of genes potentially modulated by the microbiota, 49 and 31% of the rectal and ileal DE-Gs were unique to the corresponding sites, possibly indicating specific effects of the CD altered microbiota on ileum and rectum. This is partially in line with observations from previous studies which report global differences in the transcriptional activity of genes in IBD patients between the two sites (Bogaert et al., 2010; Zhang et al., 2020b). However, these studies did not consider the plausible microbial influences and hence could be capturing a range of unknown variables on ileal and rectal gene expression. Central regulators such as SPI1, *NFKB1* and *STAT1* displayed different levels of specificity thus pointing to site-specific regulatory programs mediating the possible effects of the microbiota. Although this is a novel finding with therapeutic implications, it needs further independent validation.

By integrating different levels of information such as collection of genes associated with CD pathogenesis, targets of known CD drugs as well as gene expression signatures in response to drugs, we interpreted the disease relevance and druggability of the ileal and rectal networks. Enrichment of our networks with CD associated genes and drug targets not only assign biological significance to the inferred networks but also impart translational value. Eventhough we did not integrate mutation information from CD patients, only a third of the OpenTargets annotated CD proteins in the ileal and rectal networks are known risk loci for CD. Furthermore, expanding this to the entirety of the two signaling networks, only 6.5% on average are known IBD susceptibility loci, a subset of which include prominent components of the inflammatory pathways (NFKBIZ, STAT3, STAT4 (UniProt: Q14765), STAT5A (UniProt: P42229), STAT5B (UniProt: P51692), JAK2), chemokines (CCL2, CXCL1 (UniProt: P09341), CXCL3 (UniProt: P19876), CXCL5 (UniProt: P42830)), interleukin receptors (IL1RL1 (UniProt: Q01638), IL23R (UniProt: Q5VWK5), IL7R (UniProt: P16871)) and tight junctions (CDH1 (UniProt: Q9UM11), ITGA4, HCK (UniProt: P08631), MUC1 (UniProt: P15941)) (Table S14). This suggests that extrinsic factors such as potential microbial influences could affect the transient activity of the proteins in a mutation independent manner. Although further studies are required to ascertain this observation, it is concordant with the conclusions of many genetic studies that there is missing heritability (Chen et al., 2014; Jostins et al., 2012). From a therapeutic viewpoint, functionally and topologically relevant nodes in our networks are known targets of approved IBD drugs such as

tofacitinib (*JAK1*, *JAK2*, *JAK3*, *TYK2*) and vedolizumab (*ITGA4*).*ITGA4* interacts with vascular cell adhesion protein 1, a key protein involved in cellular adhesion and communication, and plays a major role in inflammation by recruiting memory and effector T cells to inflamed intestinal tissues(Elices et al., 1990; Reinhardt et al., 1997; Rüegg et al., 1992). In total, 24 proteins from the ileal and rectal networks could be identified cumulatively as drug targets of 27 different CD therapeutics as inferred from completed or ongoing clinical trials (Tables 2 and S11). From a translational perspective, though, topological relevant proteins in the signaling networks which are not known CD drug targets could be very handy in designing novel therapeutics since they are important mediators of plausible microbial influences on CD pathogenesis.

Despite the above-mentioned findings, our study is limited by some of the assumptions and existing bottlenecks in terms of data availability as listed below: (1) The lack of cell-type specific read-outs such as gene expression. This is relevant due to the recent studies reporting differences in cell-type compositions and activities between IBD patients and non-IBD controls (Martin et al., 2019; Smillie et al., 2019) (2) non-availability of strain-level resolution of the microbiome in both the test and validation cohorts, especially given that differences in pathogenicity are attributed to strain identities within the same species (Fang et al., 2018; Kalan et al., 2019; Zhang and Zhao, 2016) (3) lack of microbial read-outs from the intestinal mucosal samples (4) interaction prediction between microbial and host proteins are limited to predictions due to lack of large-scale bacterial-host PPIs (5) the effects of other regulatory layers such as miRNAs (Su et al., 2010) or long non-coding RNAs (Chen et al., 2013; Zacharopoulou et al., 2017) have not been considered during the network diffusion process (6) microbial components have been limited to proteins and excluded other molecules such as small RNAs, and metabolites that have been known to modulate host responses (Ahmadi Badi et al., 2020; Huang et al., 2019; Valter et al., 2020; Wlodarska et al., 2015) (7) studies have shown vast differences in composition and abundance between the rectal/colonic and ileal microbiomes (Zoetendal et al., 2008). Hence, the lack of site-specific read-outs from the ileal and rectal compartments is another drawback of the study.

To conclude, using a combinatorial computational workflow enabling the integration of high-resolution microbial datasets representing the altered gut microbiome with the gene expression from ileal and rectal biopsies of CD patients, we inferred novel mechanisms which could mediate the plausible microbial effects on host responses in CD. By performing customized analyses on the inferred signaling networks representing the signal transduction between the CD altered microbial proteins and gene expression, we identified topologically relevant proteins with functional significance as exemplified by their involvement in CD pathogenesis. Using specific examples, we pointed out potential mechanisms of action by which microbial proteins with enriched abundances or enhanced transcriptional activity can activate inflammatory responses in CD. Using gene expression data from a paired independent validation cohort of CD patients, we could validate the microbiome-modulated gene expression signatures in the ileum of CD patients. By comparing the inferred ileal and rectal signaling networks, we identified site specific differences in terms of individual genes as well as the regulatory rewiring occurring between the two sites. We also highlight functionally relevant proteins that are already known drug targets or those which could be targeted *de novo*due to their topological and functional relevance. Taken together along with the validation using an independent cohort of CD patients, our results provide deeper insights into site-specific mechanisms used by the microbiome in CD patients and provide opportunities for development of more tailored therapies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - CD patients and non-IBD control subjects
- METHOD DETAILS
  - Identifying bacterial species and genes/transcripts with enhanced abundances/transcriptional activity in CD
  - Orthology analysis

- Microbe-host protein-protein interaction prediction
- Sub-cellular localization prediction of proteins
- Gene expression in CD patients
- Inferring effects of microbe-host predictions on host gene expression
- Retrieval of biologically plausible chains
- Functional enrichment analysis
- Validation cohort
- Microbial load measurement by flow cytometry
- DNA extraction and sequencing data pre-processing
- Relative microbiome profiling (RMP)
- Quantitative microbiome profiling (QMP)
- Relative microbiome sequence variant profiling

● QUANTIFICATION AND STATISTICAL ANALYSIS

- Faecal microbiome derived features and visualization
- Microbiome and physiological features associations
- Transcriptomic analysis
- Drug and disease relevance of the networks

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.103963.

ACKNOWLEDGMENTS

SV, JS and MF are Senior Clinical Investigators of the Research Foundation Flanders [FWO], Belgium. JS is also supported by FWO and The Helmsley Charitable Trust Fund. SV is also funded by a Strategic Basic Research FWO grant [S008419N]. PS was supported by the European Research Council Advanced Grant (ERC-2015-AdG, 694679, CrUCCial) and granted to SV.

We would like to thank Evan O Paull(University of California, Santa Cruz) for providing the modified version of the TieDie algorithm. We would like to thank Dr. Dezso Modos (Quadram Institute, UK) for the valuable discussions in finetuning the equation used prior to network diffusion.

AUTHOR CONTRIBUTIONS

PS and TA performed the analysis. PS wrote the manuscript. SaV and CC provided data for the validation cohort. SaV performed the differential expression analysis for the validation cohort. MF, JS and BV provided concrete feedbacks and discussions for the manuscript and edited the manuscript. SV supervised the study and edited the manuscript. All authors read and approved the final version of the manuscript.

DECLARATION OF INTERESTS

BV reports financial support for research from Pfizer; lecture fees from Abbvie, Ferring, Takeda Pharmaceuticals, Janssen, and R Biopharm; consultancy fees from Janssen and Sandoz. JS reports lecture fees from Abbvie, Takeda, Janssen, and Nestle Health Sciences. MF reports financial support for: research from AbbVie, Amgen, Biogen, Janssen, Pfizer, Takeda; consultancy from Abbvie, Boehringer-Ingelheim, Lilly, MSD, Pfizer,Sandoz,Takeda, and Thermo Fisher; speaking from Abbvie, Amgen, Biogen, Boehringer-Ingelheim, Falk, Ferring, Janssen, Lamepro, MSD, Mylan, Pfizer, Sandoz, Takeda, and Truvion Healthcare. GM received financial support for research from DSM Nutritional Products, Karyopharm Therapeutics, and Janssen. SV reports financial support for: research from MSD, AbbVie, Takeda, Pfizer, J&J; lectures from MSD, AbbVie, Takeda, Ferring, Centocor, Hospira,Pfizer, J&J, Genentech/Roche; consultancy fromMSD,AbbVie,Takeda,Ferring, Centocor, Hospira,Pfizer, J&J,Genentech/Roche,Celgene, Mundipharma, Celltrion, Second Genome, Prometheus, Shire, Prodigest, Gilead, Galapagos.

Received: May 3, 2021
Revised: December 11, 2021
Accepted: February 18, 2022
Published: May 20, 2022

REFERENCES

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., and Nielsen, J. (2013). The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. PLoS Comput. Biol. *9*, e1002980.

Ahmadi Badi, S., Bruno, S.P., Moshiri, A., Tarashi, S., Siadat, S.D., and Masotti, A. (2020). Small RNAs in outer membrane vesicles and their function in host-microbe interactions. Front. Microbiol. *11*, 1209.

Alexander, R.J., Panja, A., Kaplan-Liss, E., Mayer, L., and Raicht, R.F. (1996). Expression of protooncogene-encoded mRNA by colonic epithelial cells in inflammatory bowel disease. Dig. Dis. Sci. *41*, 660–669.

Alpert, L., Yassan, L., Poon, R., Kadri, S., Niu, N., Patil, S.A., Mujacic, I., Montes, D., Galbo, F., Wurst, M.N., et al. (2019). Targeted mutational analysis of inflammatory bowel disease-associated colorectal cancers. Hum. Pathol. *89*, 44–50.

Altenhoff, A.M., Levy, J., Zarowiecki, M., Tomiczek, B., Warwick Vesztrocy, A., Dalquen, D.A., Müller, S., Telford, M.J., Glover, N.M., Dylus, D., et al. (2019). OMA standalone: orthology inference among public and custom genomes and transcriptomes. Genome Res. *29*, 1152–1163.

Ananthakrishnan, A.N., Luo, C., Yajnik, V., Khalili, H., Garber, J.J., Stevens, B.W., Cleland, T., and Xavier, R.J. (2017). Gut microbiome function predicts response to anti-integrin biologic therapy in inflammatory bowel diseases. Cell Host Microbe *21*, 603–610.e3.

Ananthakrishnan, A.N., Bernstein, C.N., Iliopoulos, D., Macpherson, A., Neurath, M.F., Ali, R.A.R., Vavricka, S.R., and Fiocchi, C. (2018). Environmental triggers in IBD: a review of progress and evidence. Nat. Rev. Gastroenterol. Hepatol. *15*, 39–49.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andrighetti, T., Bohar, B., Lemke, N., Sudhakar, P., and Korcsmaros, T. (2020). MicrobioLink: an integrated computational pipeline to infer functional effects of microbiome-host interactions. Cells *9*, 1278.

Arkin, A.P., Cottingham, R.W., Henry, C.S., Harris, N.L., Stevens, R.L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., et al. (2018). Kbase: the United States department of energy systems biology knowledgebase. Nat. Biotechnol. *36*, 566–569.

Atreya, R., Mudter, J., Finotto, S., Müllberg, J., Jostock, T., Wirtz, S., Schütz, M., Bartsch, B., Holtmann, M., Becker, C., et al. (2000). Blockade of interleukin 6 trans signaling suppresses T-cell resistance against apoptosis in chronic intestinal inflammation: evidence in crohn disease and experimental colitis in vivo. Nat. Med. *6*, 583–588.

Backert, I., Koralov, S.B., Wirtz, S., Kitowski, V., Billmeier, U., Martini, E., Hofmann, K., Hildner, K., Wittkopf, N., Brecht, K., et al. (2014). STAT3 activation in Th17 and Th22 cells controls IL-22-mediated epithelial host defense during infectious colitis. J. Immunol. *193*, 3779–3791.

Bengoechea, J.A., and Sa Pessoa, J. (2019). Klebsiella pneumoniae infection biology: living to counteract host defences. FEMS Microbiol. Rev. *43*, 123–144.

van Bergenhenegouwen, J., Kraneveld, A.D., Rutten, L., Garssen, J., Vos, A.P., and Hartog, A. (2016). Lipoproteins attenuate TLR2 and TLR4 activation by bacteria and bacterial ligands with differences in affinity and kinetics. BMC Immunol. *17*, 42.

Bogaert, S., Laukens, D., Peeters, H., Melis, L., Olievier, K., Boon, N., Verbruggen, G., Vandesompele, J., Elewaut, D., and De Vos, M. (2010). Differential mucosal expression of Th17-related genes between the inflamed colon and ileum of patients with inflammatory bowel disease. BMC Immunol. *11*, 61.

Bose, T., Venkatesh, K.V., and Mande, S.S. (2019). Investigating host-bacterial interactions among enteric pathogens. BMC Genomics *20*, 1022.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. Nat. Methods *13*, 581–583.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. U S A. *108*, 4516–4522.

Cario, E., and Podolsky, D.K. (2000). Differential alteration in intestinal epithelial cell expression of toll-like receptor 3 (TLR3) and TLR4 in inflammatory bowel disease. Infect. Immun. *68*, 7010–7017.

Carroll, I.M., and Maharshak, N. (2013). Enteric bacterial proteases in inflammatory bowel disease- pathophysiology and clinical implications. World J. Gastroenterol. *19*, 7531–7543.

Caruso, R., Mathes, T., Martens, E.C., Kamada, N., Nusrat, A., Inohara, N., and Núñez, G. (2019). A specific gene-microbe interaction drives the development of Crohn's disease-like colitis in mice. Sci. Immunol. *4*, eaaw4341.

Carvalho-Silva, D., Pierleoni, A., Pignatelli, M., Ong, C., Fumis, L., Karamanis, N., Carmona, M., Faulconbridge, A., Hercules, A., McAuley, E., et al. (2019). Open Targets Platform: new developments and updates two years on. Nucleic Acids Res. *47*, D1056–D1065.

Cerovic, V., Jenkins, C.D., Barnes, A.G.C., Milling, S.W.F., MacPherson, G.G., and Klavinskis, L.S. (2009). Hyporesponsiveness of intestinal dendritic cells to TLR stimulation is limited to TLR4. J. Immunol. *182*, 2405–2415.

Chen, G., Wang, Z., Wang, D., Qiu, C., Liu, M., Chen, X., Zhang, Q., Yan, G., and Cui, Q. (2013). LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res. *41*, D983–D986.

Chen, G.-B., Lee, S.H., Brion, M.-J.A., Montgomery, G.W., Wray, N.R., Radford-Smith, G.L., and Visscher, P.M.; International IBD Genetics Consortium (2014). Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. Hum. Mol. Genet. *23*, 4710–4720.

Cohavy, O., Bruckner, D., Gordon, L.K., Misra, R., Wei, B., Eggena, M.E., Targan, S.R., and Braun, J. (2000). Colonic bacteria express an ulcerative colitis pANCA-related protein epitope. Infect. Immun. *68*, 1542–1548.

Cook, L.C., Hillhouse, A.E., Myles, M.H., Lubahn, D.B., Bryda, E.C., Davis, J.W., and Franklin, C.L. (2014). The role of estrogen signaling in a mouse model of inflammatory bowel disease: a Helicobacter hepaticus model. PLoS One *9*, e94209.

Cooks, T., Pateras, I.S., Tarcic, O., Solomon, H., Schetter, A.J., Wilder, S., Lozano, G., Pikarsky, E., Forshew, T., Rosenfeld, N., et al. (2013). Mutant p53 prolongs NF-κB activation and promotes chronic inflammation and inflammation-associated colorectal cancer. Cancer Cell *23*, 634–646.

Craven, M., Egan, C.E., Dowd, S.E., McDonough, S.P., Dogan, B., Denkers, E.Y., Bowman, D., Scherl, E.J., and Simpson, K.W. (2012). Inflammation drives dysbiosis and bacterial invasion in murine models of ileal Crohn's disease. PLoS One *7*, e41594.

Cui, T., Li, W., Liu, L., Huang, Q., and He, Z.-G. (2016). Uncovering newpathogen-host protein-protein interactions by pairwise structure similarity. PLoS One *11*, e0147612.

Dalwadi, H., Wei, B., Kronenberg, M., Sutton, C.L., and Braun, J. (2001). The Crohn's disease-associated bacterial protein I2 is a novel enteric t cell superantigen. Immunity *15*, 149–158.

De Cruz, P., Kang, S., Wagner, J., Buckley, M., Sim, W.H., Prideaux, L., Lockett, T., McSweeney, C., Morrison, M., Kirkwood, C.D., et al. (2015). Association between specific mucosa-associated microbiota in Crohn's disease at the time of resection and subsequent disease recurrence: a pilot study. J. Gastroenterol. Hepatol. *30*, 268–278.

Delignette-Muller, M.-L. (2021). Fitdistrplus: Help to Fit of a Parametric Distribution to Non-censored or Censored Data (Comprehensive R Archive Network (CRAN)).

Derek, O. (2021). FSA Package (Comprehensive R Archive Network (CRAN)).

Dinkel, H., Van Roey, K., Michael, S., Davey, N.E., Weatheritt, R.J., Born, D., Speck, T., Krüger, D., Grebnev, G., Kuban, M., et al. (2014). The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res. *42*, D259–D266.

Doherty, M.K., Ding, T., Koumpouras, C., Telesco, S.E., Monast, C., Das, A., Brodmerkel, C., and Schloss, P.D. (2018). Fecal microbiota signatures are associated with response to ustekinumab therapy among crohn's disease patients. MBio. *9*, e02120–17.

Duan, Q., Reid, S.P., Clark, N.R., Wang, Z., Fernandez, N.F., Rouillard, A.D., Readhead, B., Tritsch, S.R., Hodos, R., Hafner, M., et al. (2016). L1000CDS2: LINCS L1000 characteristic direction signatures search engine. NPJ Syst. Biol. Appl. *2*, 16015.

Durant, L., Watford, W.T., Ramos, H.L., Laurence, A., Vahedi, G., Wei, L., Takahashi, H., Sun, H.-W., Kanno, Y., Powrie, F., et al. (2010). Diverse targets of the transcription factor STAT3 contribute to T cell pathogenicity and homeostasis. Immunity *32*, 605–615.

Durmuş Tekir, S., Çakır, T., Ardiç, E., Sayılırbaş, A.S., Konuk, G., Konuk, M., Sarıyer, H., Uğurlu, A., Karadeniz, İ., Özgür, A., et al. (2013). PHISTO: pathogen-host interaction search tool. Bioinformatics *29*, 1357–1358.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics *27*, 2194–2200.

Elices, M.J., Osborn, L., Takada, Y., Crouse, C., Luhowskyj, S., Hemler, M.E., and Lobb, R.R. (1990). VCAM-1 on activated endothelium interacts with the leukocyte integrin VLA-4 at a site distinct from the VLA-4/fibronectin binding site. Cell *60*, 577–584.

Elsayed, S., and Zhang, K. (2004). Human infection caused by Clostridium hathewayi. Emerging Infect. Dis. *10*, 1950–1952.

Erickson, A.R., Cantarel, B.L., Lamendella, R., Darzi, Y., Mongodin, E.F., Pan, C., Shah, M., Halfvarson, J., Tysk, C., Henrissat, B., et al. (2012). Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. PLoS One *7*, e49138.

Fagnart, O.C., Mareschal, J.C., Cambiaso, C.L., and Masson, P.L. (1985). Particle-counting immunoassay (PACIA) of pregnancy-specific beta 1-glycoprotein, a possible marker of various malignancies and Crohn's ileitis. Clin. Chem. *31*, 397–401.

Fang, X., Monk, J.M., Nurk, S., Akseshina, M., Zhu, Q., Gemmell, C., Gianetto-Hill, C., Leung, N., Szubin, R., Sanders, J., et al. (2018). Metagenomics-based, strain-level analysis of Escherichia coli from a time-series of microbiome samples from a crohn's disease patient. Front. Microbiol. *9*, 2559.

Ferguson, L.R., Han, D.Y., Fraser, A.G., Huebner, C., Lam, W.J., Morgan, A.R., Duan, H., and Karunasinghe, N. (2010). 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. *690*, 108–115.

Franzosa, E.A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H.J., Reinker, S., Vatanen, T., Hall, A.B., Mallick, H., McIver, L.J., et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat. Microbiol. *4*, 293–305.

Gao, Z., Chen, S., Qin, S., and Tang, C. (2018). Network motifs capable of decoding transcription factor dynamics. Sci. Rep. *8*, 3594.

Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. *29*, 1363–1375.

Gevers, D., Kugathasan, S., Denson, L.A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., et al. (2014). The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe *15*, 382–392.

Goldsmith, J.R., Uronis, J.M., and Jobin, C. (2011). Mu opioid signaling protects against acute murine intestinal injury in a manner involving Stat3 signaling. Am. J. Pathol. *179*, 673–683.

Goodman, W.A., Bedoyan, S.M., Havran, H.L., Richardson, B., Cameron, M.J., and Pizarro, T.T. (2020). Impaired estrogen signaling underlies regulatory T cell loss-of-function in the chronically inflamed intestine. Proc. Natl. Acad. Sci. U S A. *117*, 17166–17176.

Gordon, D.E., Hiatt, J., Bouhaddou, M., Rezelj, V.V., Ulferts, S., Braberg, H., Jureka, A.S., Obernier, K., Guo, J.Z., Batra, J., et al. (2020). Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. Science *370*, eabe9403.

Gouw, M., Michael, S., Sámano-Sánchez, H., Kumar, M., Zeke, A., Lang, B., Bely, B., Chemes, L.B., Davey, N.E., Deng, Z., et al. (2018). The eukaryotic linear motif resource - 2018 update. Nucleic Acids Res. *46*, D428–D434.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

Guven-Maiorov, E., Tsai, C.-J., and Nussinov, R. (2017). Structural host-microbiota interaction networks. PLoS Comput. Biol. *13*, e1005579.

Halfvarson, J., Brislawn, C.J., Lamendella, R., Vázquez-Baeza, Y., Walters, W.A., Bramer, L.M., D'Amato, M., Bonfiglio, F., McDonald, D., Gonzalez, A., et al. (2017). Dynamics of the human gut microbiome in inflammatory bowel disease. Nat. Microbiol. *2*, 17004.

Han, Y.M., Koh, J., Kim, J.W., Lee, C., Koh, S.-J., Kim, B., Lee, K.L., Im, J.P., and Kim, J.S. (2017). NF-kappa B activation correlates with disease phenotype in Crohn's disease. PLoS One *12*, e0182071.

Hartman, D.J., Binion, D.G., Regueiro, M.D., Miller, C., Herbst, C., and Pai, R.K. (2018). Distinct histopathologic and molecular alterations in inflammatory bowel disease-associated intestinal adenocarcinoma: c-MYC amplification is common and associated with mucinous/signet ring cell differentiation. Inflamm. Bowel Dis. *24*, 1780–1790.

Hausmann, M., Kiessling, S., Mestermann, S., Webb, G., Spöttl, T., Andus, T., Schölmerich, J., Herfarth, H., Ray, K., Falk, W., et al. (2002). Toll-like receptors 2 and 4 are up-regulated during intestinal inflammation. Gastroenterology *122*, 1987–2000.

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., et al. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat. Protoc. *14*, 639–702.

Holmes, I., Harris, K., and Quince, C. (2012). Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS One *7*, e30126.

Hothorn, T. (2021). R Package "Coin (Comprehensive R Archive Network (CRAN)).

Huang, J., and Brumell, J.H. (2014). Bacteria-autophagy interplay: a battle for survival. Nat. Rev. Microbiol. *12*, 101–114.

Huang, C.-Y., Wang, H., Hu, P., Hamby, R., and Jin, H. (2019). Small RNAs - big players in plant-microbe interactions. Cell Host Microbe *26*, 173–182.

Johansson, M.E.V., Gustafsson, J.K., Holmén-Larsson, J., Jabbar, K.S., Xia, L., Xu, H., Ghishan, F.K., Carvalho, F.A., Gewirtz, A.T., Sjövall, H., et al. (2014). Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. Gut *63*, 281–291.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature *491*, 119–124.

Kalan, L.R., Meisel, J.S., Loesche, M.A., Horwinski, J., Soaita, I., Chen, X., Uberoi, A., Gardner, S.E., and Grice, E.A. (2019). Strain- and species-level variation in the microbiome of diabetic wounds is associated with clinical outcomes and therapeutic efficacy. Cell Host Microbe *25*, 641–655.e5.

Kanehisa, M., Sato, Y., and Morishima, K. (2016). Blastkoala and ghostkoala: KEGG tools for functional characterization of genome and metagenome sequences. J. Mol. Biol. *428*, 726–731.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. *45*, D353–D361.

Karin, M., and Smeal, T. (1992). Control of transcription factors by signal transduction pathways: the beginning of the end. Trends Biochem. Sci. *17*, 418–422.

Kekuda, R., Saha, P., and Sundaram, U. (2008). Role of Sp1 and HNF1 transcription factors in SGLT1 regulation during chronic intestinal inflammation. Am. J. Physiol. Gastrointest. Liver Physiol. *294*, G1354–G1361.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kim, J.-K., Lee, S.H., Lee, S.-Y., Kim, E.-K., Kwon, J.-E., Seo, H.-B., Lee, H.H., Lee, B.-I., Park, S.-H., and Cho, M.-L. (2016). Grim19 attenuates DSS induced colitis in an animal model. PLoS One *11*, e0155853.

Kincaid, R.K., and Phillips, D.J. (2011). Network topology measures. Wires Comp. Stat. *3*, 557–565.

Kløve, S., Genger, C., Mousavi, S., Weschka, D., Bereswill, S., and Heimesaat, M.M. (2020). Toll-likereceptor-4 dependent intestinal and systemic sequelae following peroral Campylobacter coli infection of IL10 deficient mice harboring a human gut microbiota. Pathogens *9*, 386.

Knapp, S., Brodal, C., Peterson, J., Qi, F., Kreth, J., and Merritt, J. (2017). Natural competence is common among clinical isolates of Veillonella parvula and is useful for genetic manipulation of this key member of the oral microbiome. Front. Cell Infect. Microbiol. *7*, 139.

Koay, L.C., Rigby, R.J., and Wright, K.L. (2014). Cannabinoid-induced autophagy regulates suppressor of cytokine signaling-3 in intestinal epithelium. Am. J. Physiol. Gastrointest. Liver Physiol. *307*, G140–G148.

Koboziev, I., Reinoso Webb, C., Furr, K.L., and Grisham, M.B. (2014). Role of the enteric microbiota in intestinal homeostasis and inflammation. Free Radic. Biol. Med. *68*, 122–133.

Krishnadev, O., and Srinivasan, N. (2011). Prediction of protein-protein interactions between human host and a pathogen and its application to three pathogenic bacteria. Int. J. Biol. Macromol. *48*, 613–619.

Kshirsagar, M., Carbonell, J., and Klein-Seetharaman, J. (2012). Techniques to cope with missing data in host-pathogen protein interaction prediction. Bioinformatics *28*, i466–i472.

Kumar, R., and Nanduri, B. (2010). HPIDB–a unified resource for host-pathogen interactions. BMC Bioinformatics *11*, S16.

Laurent, C., Svrcek, M., Flejou, J.-F., Chenard, M.-P., Duclos, B., Freund, J.-N., and Reimund, J.-M. (2011). Immunohistochemical expression of CDX2, β-catenin, and TP53 in inflammatory bowel disease-associated colorectal cancer. Inflamm. Bowel Dis. *17*, 232–240.

Leon-Cabrera, S., Vázquez-Sandoval, A., Molina-Guzman, E., Delgado-Ramirez, Y., Delgado-Buenrostro, N.L., Callejas, B.E., Chirino, Y.I., Pérez-Plasencia, C., Rodríguez-Sosa, M., Olguín, J.E., et al. (2018). Deficiency in STAT1 signaling predisposes gut inflammation and prompts colorectal cancer development. Cancers (Basel) *10*, 341.

Levy, R., Carr, R., Kreimer, A., Freilich, S., and Borenstein, E. (2015). NetCooperate: a network-based tool for inferring host-microbe and microbe-microbe cooperation. BMC Bioinformatics *16*, 164.

Li, X., Zhang, Z., Li, L., Gong, W., Lazenby, A.J., Swanson, B.J., Herring, L.E., Asara, J.M., Singer, J.D., and Wen, H. (2017). Myeloid-derived cullin 3 promotes STAT3 phosphorylation by inhibiting OGT expression and protects against intestinal inflammation. J. Exp. Med. *214*, 1093–1109.

Li, Y., Wang, C., Miao, Z., Bi, X., Wu, D., Jin, N., Wang, L., Wu, H., Qian, K., Li, C., et al. (2015). ViRBase: a resource for virus-host ncRNA-associated interactions. Nucleic Acids Res. *43*, D578–D582.

Liu, X., Lee, Y.S., Yu, C.-R., and Egwuagu, C.E. (2008). Loss of STAT3 in CD4+ T cells prevents development of experimental autoimmune diseases. J. Immunol. *180*, 6070–6076.

Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N., Schirmer, M., Avila-Pacheco, J., Poon, T.W., Andrews, E., Ajami, N.J., Bonham, K.S., Brislawn, C.J., et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature *569*, 655–662.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963.

Martin, J.C., Chang, C., Boschetti, G., Ungaro, R., Giri, M., Grout, J.A., Gettler, K., Chuang, L.-S., Nayar, S., Greenstein, A.J., et al. (2019). Single-cell analysis of crohn's disease lesions identifies a pathogenic cellular module associated with resistance to anti-TNF therapy. Cell *178*, 1493–1508.e20.

McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One *8*, e61217.

Mészáros, B., Erdos, G., and Dosztányi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. *46*, W329–W337.

Metwaly, A., Dunkel, A., Waldschmitt, N., Raj, A.C.D., Lagkouvardos, I., Corraliza, A.M., Mayorgas, A., Martinez-Medina, M., Reiter, S., Schloter, M., et al. (2020). Integrated microbiota and metabolite profiles link Crohn's disease to sulfur metabolism. Nat. Commun. *11*, 4322.

Monteleone, G., Kumberova, A., Croft, N.M., McKenzie, C., Steer, H.W., and MacDonald, T.T. (2001). Blocking Smad7 restores TGF-beta1 signaling in chronic inflammatory bowel disease. J. Clin. Invest. *108*, 601–609.

Morgan, M. (2021). DirichletMultinomial: Dirichlet-Multinomial Mixture Model Machine Learning for Microbiome Data. R Package Version 1.32.0 (Comprehensive R Archive Network(CRAN)).

Mudter, J., Weigmann, B., Bartsch, B., Kiesslich, R., Strand, D., Galle, P.R., Lehr, H.A., Schmidt, J., and Neurath, M.F. (2005). Activation pattern of signal transducers and activators of transcription (STAT) factors in inflammatory bowel diseases. Am. J. Gastroenterol. *100*, 64–72.

Musso, A., Dentelli, P., Carlino, A., Chiusa, L., Repici, A., Sturm, A., Fiocchi, C., Rizzetto, M., Pegoraro, L., Sategna-Guidetti, C., et al. (2005). Signal transducers and activators of transcription 3 signaling pathway: an essential mediator of inflammatory bowel disease and other forms of intestinal inflammation. Inflamm. Bowel Dis. *11*, 91–98.

Naftali, T., Reshef, L., Kovacs, A., Porat, R., Amir, I., Konikoff, F.M., and Gophna, U. (2016). Distinct microbiotas are associated with ileum-restricted and colon-involving crohn's disease. Inflamm. Bowel Dis. *22*, 293–302.

Nourani, E., Khunjush, F., and Durmuş, S. (2015). Computational approaches for prediction of pathogen-host protein-protein interactions. Front. Microbiol. *6*, 94.

Oksanen, J. (2021). CRAN - Package Vegan (Comprehensive R Archive Network (CRAN)).

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., Martinez, X., Varela, E., Sarrabayrouse, G., Machiels, K., et al. (2017). A microbial signature for Crohn's disease. Gut *66*, 813–822.

Paull, E.O., Carlin, D.E., Niepel, M., Sorger, P.K., Haussler, D., and Stuart, J.M. (2013). Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion through Interacting Events (TieDIE). Bioinformatics *29*, 2757–2764.

Pérez-Brocal, V., García-López, R., Nos, P., Beltrán, B., Moret, I., and Moya, A. (2015). Metagenomic analysis of crohn's disease patients identifies changes in the virome and microbiome related to disease status and therapy, and detects potential interactions and biomarkers. Inflamm. Bowel Dis. *21*, 2515–2532.

Ramakrishnan, S.K., Zhang, H., Ma, X., Jung, I., Schwartz, A.J., Triner, D., Devenport, S.N., Das, N.K., Xue, X., Zeng, M.Y., et al. (2019). Intestinal non-canonical NFκB signaling shapes the local and systemic immune response. Nat. Commun. *10*, 660.

Rastogi, S., and Rost, B. (2011). LocDB: experimental annotations of localization for Homo sapiens and Arabidopsis thaliana. Nucleic Acids Res. *39*, D230–D234.

Reinhardt, P.H., Elliott, J.F., and Kubes, P. (1997). Neutrophils can adhere via alpha4beta1-integrin under flow conditions. Blood *89*, 3837–3846.

Ritz, A., Poirel, C.L., Tegge, A.N., Sharp, N., Simmons, K., Powell, A., Kale, S.D., and Murali, T.M. (2016). Pathways on demand: automated reconstruction of human signaling networks. NPJ Syst. Biol. Appl. *2*, 16002.

Roberton, A.M., and Corfield, A.P. (1999). Mucin degradation and its significance in inflammatory conditions of the gastrointestinal tract. In Medical Importance of the Normal Microflora, G.W. Tannock, ed. (Springer US), pp. 222–261.

Rüegg, C., Postigo, A.A., Sikorski, E.E., Butcher, E.C., Pytela, R., and Erle, D.J. (1992). Role of integrin alpha 4 beta 7/alpha 4 beta P in lymphocyte adherence to fibronectin and VCAM-1 and in homotypic cell clustering. J. Cell Biol. *117*, 179–189.

Savić Mlakar, A., Hojsak, I., Jergović, M., Čimić, S., and Bendelja, K. (2018). Pediatric Crohn disease is characterized by Th1 in the terminal ileum and Th1/Th17 immune response in the colon. Eur. J. Pediatr. *177*, 611–616.

Schaubeck, M., Clavel, T., Calasan, J., Lagkouvardos, I., Haange, S.B., Jehmlich, N., Basic, M., Dupont, A., Hornef, M., von Bergen, M., et al. (2016). Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis

independent of failure in antimicrobial defence. Gut *65*, 225–237.

Schreiber, S., Nikolaus, S., and Hampe, J. (1998). Activation of nuclear factor kappa B inflammatory bowel disease. Gut *42*, 477–484.

Schweppe, D.K., Harding, C., Chavez, J.D., Wu, X., Ramage, E., Singh, P.K., Manoil, C., and Bruce, J.E. (2015). Host-microbe protein interactions during bacterial infection. Chem. Biol. *22*, 1521–1530.

Segain, J.P., Raingeard de la Blétière, D., Bourreille, A., Leray, V., Gervois, N., Rosales, C., Ferrier, L., Bonnet, C., Blottière, H.M., and Galmiche, J.P. (2000). Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease. Gut *47*, 397–403.

Shah, M., Rennoll, S.A., Raup-Konsavage, W.M., and Yochum, G.S. (2015). A dynamic exchange of TCF3 and TCF4 transcription factors controls MYC expression in colorectal cancer cells. Cell Cycle *14*, 323–332.

Shaw, K.A., Bertha, M., Hofmekler, T., Chopra, P., Vatanen, T., Srivatsa, A., Prince, J., Kumar, A., Sauer, C., Zwick, M.E., et al. (2016). Dysbiosis, inflammation, and response to treatment: a longitudinal study of pediatric subjects with newly diagnosed inflammatory bowel disease. Genome Med. *8*, 75.

Silva, M.A., Quera, R., Valenzuela, J., Salim, S.Y., Söderholm, J.D., and Perdue, M.H. (2008). Dendritic cells and toll-like receptors 2 and 4 in the ileum of Crohn's disease patients. Dig. Dis. Sci. *53*, 1917–1928.

Smillie, C.S., Biton, M., Ordovas-Montanes, J., Sullivan, K.M., Burgin, G., Graham, D.B., Herbst, R.H., Rogel, N., Slyper, M., Waldman, J., et al. (2019). Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. Cell *178*, 714–730.e22.

Sorrentino, D. (2017). Noncanonical NF-κB signaling is elevated in inflammatory bowel disease patients and may be associated with therapeutic response. J.Immunol. *198*, 197.5.

Spehlmann, M.E., Manthey, C.F., Dann, S.M., Hanson, E., Sandhu, S.S., Liu, L.Y., Abdelmalak, F.K., Diamanti, M.A., Retzlaff, K., Scheller, J., et al. (2013). Trp53 deficiency protects against acute intestinal inflammation. J. Immunol. *191*, 837–847.

Stoddard, S.F., Smith, B.J., Hein, R., Roller, B.R.K., and Schmidt, T.M. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res. *43*, D593–D598.

Su, N., Wang, Y., Qian, M., and Deng, M. (2010). Combinatorial regulation of transcription factors and microRNAs. BMC Syst. Biol. *4*, 150.

Sudhakar, P., Jacomin, A.-C., Hautefort, I., Samavedam, S., Fatemian, K., Ari, E., Gul, L., Demeter, A., Jones, E., Korcsmaros, T., et al. (2019). Targeted interplay between bacterial pathogens and host autophagy. Autophagy *15*, 1620–1633.

Suzuki, A., Hanada, T., Mitsuyama, K., Yoshida, T., Kamizono, S., Hoshino, T., Kubo, M., Yamashita, A., Okabe, M., Takeda, K., et al. (2001). CIS3/SOCS3/SSI3 plays a negative regulatory role in STAT3 activation and intestinal inflammation. J. Exp. Med. *193*, 471–481.

Ta, A., Thakur, B.K., Dutta, P., Sinha, R., Koley, H., and Das, S. (2017). Double-stranded RNA induces cathelicidin expression in the intestinal epithelial cells through phosphatidylinositol 3-kinase-protein kinase Cζ-Sp1 pathway and ameliorates shigellosis in mice. Cell Signal. *35*, 140–153.

Taniguchi, K., Wu, L.-W., Grivennikov, S.I., de Jong, P.R., Lian, I., Yu, F.-X., Wang, K., Ho, S.B., Boland, B.S., Chang, J.T., et al. (2015). A gp130-Src-YAP module links inflammation to epithelial regeneration. Nature *519*, 57–62.

The UniProt Consortium (2018). UniProt: the universal protein knowledgebase. Nucleic Acids Res. *46*, 2699.

Thul, P.J., and Lindskog, C. (2018). The human protein atlas: a spatial map of the human proteome. Protein Sci. *27*, 233–244.

Türei, D., Korcsmáros, T., and Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat. Methods *13*, 966–967.

Valter, M., Verstockt, S., Finalet Ferreiro, J.A., and Cleynen, I. (2020). Extracellular vesicles in inflammatory bowel disease: small particles, big players. J. Crohns Colitis *15*, 499–510.

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R.Y., De Commer, L., Darzi, Y., et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. Nature *551*, 507–511.

Verstockt, B., Verstockt, S., Veny, M., Dehairs, J., Arnauts, K., Van Assche, G., De Hertogh, G., Vermeire, S., Salas, A., and Ferrante, M. (2019). Expression levels of 4 genes in colon tissue might be used to predict which patients will enter endoscopic remission after vedolizumab therapy for inflammatory bowel diseases. Clin.Gastroenterol. Hepatol. *18*, 1142–1151.

Vialas, V., and Gil, C. (2015). Proteopathogen2, a database and web tool to store and display proteomics identification results in the mzIdentML standard. EuPA Open Proteom. *8*, 22–27.

Wanders, L.K., Cordes, M., Voorham, Q., Sie, D., de Vries, S.D., d'Haens, G.R.A.M., de Boer, N.K.H., Ylstra, B., van Grieken, N.C.T., Meijer, G.A., et al. (2020). IBD-associated dysplastic lesions show more chromosomal instability thansporadic adenomas. Inflamm. Bowel Dis. *26*, 167–180.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. *73*, 5261–5267.

Willson, T.A., Kuhn, B.R., Jurickova, I., Gerad, S., Moon, D., Bonkowski, E., Carey, R., Collins, M.H., Xu, H., Jegga, A.G., et al. (2012). STAT3 genotypic variation and cellular STAT3 activation and colon leukocyte recruitment in pediatric Crohn disease. J. Pediatr. Gastroenterol.Nutr. *55*, 32–43.

Wlodarska, M., Kostic, A.D., and Xavier, R.J. (2015). An integrative view of microbiome-host interactions in inflammatory bowel diseases. Cell Host Microbe *17*, 577–591.

Wu, F., Dassopoulos, T., Cope, L., Maitra, A., Brant, S.R., Harris, M.L., Bayless, T.M., Parmigiani, G., and Chakravarti, S. (2007). Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: insights into distinctive pathogenesis. Inflamm. Bowel Dis. *13*, 807–821.

Wu, L.-L., Chiu, H.-D., Peng, W.-H., Lin, B.-R., Lu, K.-S., Lu, Y.-Z., and Yu, L.C.-H. (2011). Epithelial inducible nitric oxide synthase causes bacterial translocation by impairment of enterocytic tight junctions via intracellular signals of Rho-associated kinase and protein kinase C zeta. Crit. Care Med. *39*, 2087–2098.

Wu, L.-L., Peng, W.-H., Kuo, W.-T., Huang, C.-Y., Ni, Y.-H., Lu, K.-S., Turner, J.R., and Yu, L.C.H. (2014). Commensal bacterial endocytosis in epithelial cells is dependent on myosin light chain kinase-activated brush border fanning by interferon-γ. Am. J. Pathol. *184*, 2260–2274.

Yang, X., Letterio, J.J., Lechleider, R.J., Chen, L., Hayman, R., Gu, H., Roberts, A.B., and Deng, C. (1999). Targeted disruption of SMAD3 results in impaired mucosal immunity and diminished T cell responsiveness to TGF-beta. EMBO J. *18*, 1280–1291.

Yu, L.C.-H. (2015). Commensal bacterial internalization by epithelial cells: an alternative portal for gut leakiness. Tissue Barriers *3*, e1008895.

Yu, G., and He, Q.-Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol. Biosyst. *12*, 477–479.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

Yu, L.C.-H., Shih, Y.-A., Wu, L.-L., Lin, Y.-D., Kuo, W.-T., Peng, W.-H., Lu, K.-S., Wei, S.-C., Turner, J.R., and Ni, Y.-H. (2014). Enteric dysbiosis promotes antibiotic-resistant bacterial infection: systemic dissemination of resistant and commensal bacteria through epithelial transcytosis. Am. J. Physiol. Gastrointest. Liver Physiol. *307*, G824–G835.

Zacharopoulou, E., Gazouli, M., Tzouvala, M., Vezakis, A., and Karamanolis, G. (2017). The contribution of long non-coding RNAs in Inflammatory Bowel Diseases. Dig. Liver Dis. *49*, 1067–1072.

Zhang, C., and Zhao, L. (2016). Strain-level dissection of the contribution of the gut microbiome to human metabolic disease. Genome Med. *8*, 41.

Zhang, X., Deeke, S.A., Ning, Z., Starr, A.E., Butcher, J., Li, J., Mayne, J., Cheng, K., Liao, B., Li, L., et al. (2018). Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric

inflammatory bowel disease. Nat. Commun. *9*, 2873.

Zhang, X., Ning, Z., Mayne, J., Yang, Y., Deeke, S.A., Walker, K., Farnsworth, C.L., Stokes, M.P., Couture, J.-F., Mack, D., et al. (2020a). Widespread protein lysine acetylation in gut microbiome and its alterations in patients with Crohn's disease. Nat. Commun. *11*, 4120.

Zhang, Y., Shen, B., Zhuge, L., and Xie, Y. (2020b). Identification of differentially expressed genes between the colon and ileum of patients with inflammatory bowel disease by gene co-expression analysis. J. Int. Med. Res. *48*, 300060519887268.

Zoetendal, E.G., Rajilic-Stojanovic, M., and de Vos, W.M. (2008). High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. Gut *57*, 1605–1615.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Ileal tissue biopsies of Crohn's disease patients and non-IBD control subjects	UZ Leuven/KU Leuven biobank	https://www.uzleuven.be/en/uz-kuleuven-biobank
Fecal samples of Crohn's disease patients and non-IBD control subjects	UZ Leuven/KU Leuven biobank	https://www.uzleuven.be/en/uz-kuleuven-biobank
Chemicals, peptides, and recombinant proteins		
SYBR Green I	Thermo Fisher Scientific, Massachusetts, USA	Cat#S7563
Critical commercial assays		
PowerMicrobiome RNA isolation kit	Mo Bio	Cat#26000-50
Deposited data		
Raw fastq files	This paper	E-MTAB-10395
Software and algorithms		
BlastKOALA	(Kanehisa et al., 2016)	https://www.kegg.jp/blastkoala/
MicrobioLink	(Andrighetti et al., 2020)	https://github.com/korcsmarosgroup/HMIpipeline
IUpred2A	(Mészáros et al., 2018)	http://iupred2a.elte.hu./
TieDie	(Paull et al., 2013)	https://github.com/epaull/TieDIE
PathLinker	(Ritz et al., 2016)	https://github.com/Murali-group/PathLinker
ReactomePA	(Yu and He, 2016)	http://bioconductor.org/packages/release/bioc/html/ReactomePA.html
clusterProfiler	(Yu et al., 2012)	http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
FLASH	(Magoč and Salzberg, 2011)	http://www.cbcb.umd.edu/software/flash
UCHIME	(Edgar et al., 2011)	http://drive5.com/usearch/manual/uchime_algo.html
Fastx tool kit	Hannon Lab	http://hannonlab.cshl.edu/fastx_toolkit/
DADA2 pipeline v1.6.0	(Naftali et al., 2016)	https://benjjneb.github.io/dada2/tutorial.html
RDP classifier 2.12	(Wang et al., 2007)	https://bioweb.pasteur.fr/packages/pack@rdp_classifier@2.12
rrnDB	(Stoddard et al., 2015)	https://rrndb.umms.med.umich.edu/
PhyML	(Guindon et al., 2010)	http://www.atgc-montpellier.fr/phyml/
Vegan	(Oksanen, 2021)	https://cran.r-project.org/web/packages/vegan/index.html
Phyloseq	(McMurdie and Holmes, 2013)	https://joey711.github.io/phyloseq/
FSA	(Derek, 2021)	https://cran.r-project.org/web/packages/FSA/index.html
Coin	(Hothorn, 2021)	https://cran.r-project.org/web/packages/coin/index.html
DirichletMultinomial	(Morgan, 2021)	http://bioconductor.org/packages/release//bioc/html/DirichletMultinomial.html
fitdistrplus	(Delignette-Muller, 2021)	https://cran.r-project.org/web/packages/fitdistrplus/index.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Hisat2 version 2.1.0	(Kim et al., 2015)	http://daehwankimlab.github.io/hisat2/manual/
HTSeq	(Anders et al., 2015)	https://htseq.readthedocs.io/en/master/
DESeq2 package	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
BD Accuri CFlow software	BDbiosciences	https://www.bdbiosciences.com/us/instruments/research/cell-analyzers/bd-accuri/software/c6-plus-analysis-software-for-pc-or-mac/p/661083
Other		
IBDMDB	(Lloyd-Price et al., 2019)	https://ibdmdb.org/tunnel/public/summary.html
Uniprot	(The UniProt Consortium, 2018)	https://www.uniprot.org/
DoRothEA	(Garcia-Alonso et al., 2019)	https://github.com/saezlab/dorothea-py
OmniPath	(Türei et al., 2016)	https://omnipathdb.org/
LocDB	(Rastogi and Rost, 2011)	https://www.rostlab.org/services/locDB/
Human Protein Atlas (HPA)	(Thul and Lindskog, 2018)	https://www.proteinatlas.org/
OpenTargets	(Carvalho-Silva et al., 2019)	https://www.targetvalidation.org/
LINCS L1000 dataset	(Duan et al., 2016)	https://lincsproject.org/LINCS/tools/workflows/find-the-best-place-to-obtain-the-lincs-l1000-data
IBD risk loci	(Jostins et al., 2012)	PMID: 23128233

RESOURCE AVAILABILITY

Lead contact

Further information and requests for reagents should be directed to Dr. Padhmanand Sudhakar (padhmanand.sudhakar@kuleuven.be).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The raw datasets used for the study have been deposited at ArrayExpress (accession number: E-MTAB-10395).
- All the original and generic codes are available via open access tools and resources listed in the STAR Methods and key resources table.
- Additional information supporting the findings will be made available upon request from the lead contact Dr. Padhmanand Sudhakar (padhmanand.sudhakar@kuleuven.be).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

CD patients and non-IBD control subjects

For the validation cohort, patients were selected based on (a) their identification as having active CD as assessed by the endoscopic scores and (b) having paired fecal microbiome profiling datasets. Non-IBD controls were similarly chosen (paired ileal gene expression and fecal microbiome profiling) from the Flemish Gut Flora Project. The ethics committee of the University Hospitals Leuven approved the study (IRB approvals, B322201213950/S53684 and B322201627472/S57662). All individuals gave written informed consent.

METHOD DETAILS

Identifying bacterial species and genes/transcripts with enhanced abundances/ transcriptional activity in CD

Bacterial species relevant for CD were selected (Table S1) based on multiple lines of evidence. Firstly, from the IBD Multi-omics Database within the Human Integrated Microbiome project (Lloyd-Price et al., 2019), bacterial species with alterations between CD (n = 67) and non-IBD subjects (n = 27) at both the levels of abundance and transcriptional activity, species with enhanced abundances displaying a t-statistic of ≥ 2 were selected. From the same study, bacterial species contributing to alterations in the community composition as measured by the dysbiotic index between CD and non-IBD subjects (at both the levels of abundance and transcriptional activity) displaying enhanced abundances displaying a t-statistic of ≥ 2 and Q-value ≤ 0.05 were selected. The bacterial species identified above were subjected to a literature search to check if they have been observed in at least one other study to have increased abundance levels in CD patients. This was done to improve the reliability of our inferences about the bacterial species inferred as being important in the context of CD (Table S1). Genes with enhanced abundances/transcriptional activity in CD were determined from the metagenomic and metatranscriptomic profiles. Significant metagenomic Enzyme Classification (EC) profiles, EC transcription, metagenomic KEGG Orthology (KO) profiles, KO transcription, protein abundances, EC-level protein abundances, and KO-level protein abundances with enhanced abundances in CD from the HMP2 study(Lloyd-Price et al., 2019) were identified by setting a threshold for the t-statistic at ≥ 2 and Q-value ≤ 0.05 (Table S2). The KO-number (KEGG orthology number) of the genes/transcripts with significant abundances were retrieved from KEGG (Kanehisa et al., 2017). De-novoprediction of K-numbers for the chosen proteomes were performed using BlastKOALA (Kanehisa et al., 2016).

Orthology analysis

The annotated proteomes corresponding to the identified bacterial species were downloaded from Uniprot (The UniProt Consortium, 2018). Where available, the reference proteomes were retrieved. The stand-alone version of the ortholog group mapping software OMA (Altenhoff et al., 2019) was used to identify orthologous groups. Default settings were used to carry out the ortholog analysis. Orthology tests against multiple members of Faecalibacterium sp. and Akkarmensia sp were performed using the NCBI blastp with default parameter settings (E-value ≤ 0.05).

Microbe-host protein-protein interaction prediction

The microbe-host interaction prediction based on the inferred interactions and their downstream effects was carried out along the lines of the general framework laid out in the manuscript by Andrighetti et al. (2020). To maintain the context specificity (i.e. prevalence in CD), interactions corresponding to bacterial genes/proteins with enhanced abundances in CD (as against the non-IBD controls) (Table S2) were considered for further analysis. Interaction prediction between microbial and host proteins were performed by using the domain-domain and domain-motif interaction prediction principle applied in the manuscript by Andrighetti et al. (2020) and used elsewhere in previous studies (Sudhakar et al., 2019). Briefly, the interaction prediction assumes that protein-pairs with known interacting domain-motif or domain-domain pairs also interact. With the domain-motif method, in order to capture the desired direction of signal transduction (i.e., the modulation of host proteins by the bacterial proteins), we imposed the condition that the interacting domains need to be harbored by the microbial proteins and the motifs on the host proteins. Pfam domain annotation for the microbial proteins was retrieved from Uniprot (The UniProt Consortium, 2018) whereas information about the occurrence of motifs and the gold standard list of domain-motif interactions were obtained from the Eukaryotic Linear Motif database (Dinkel et al., 2014; Gouw et al., 2018). The resulting set of predicted interactions between microbial and host proteins were then subjected to a disordered region based structural filtering procedure as described by Andrighetti et al. (2020). Disordered region predictions were performed by using IUpred2A (Mészáros et al., 2018). In short, only protein-protein interactions corresponding to motifs that lie within disordered regions of the host proteins were considered.

Sub-cellular localization prediction of proteins

Where available, the reference proteomes corresponding to the bacterial species were downloaded from Uniprot (The UniProt Consortium, 2018).Since we did not find evidence for the majority of inferred bacterial species to be intracellular (with the exception of Adherent-Invasive *E. coli*(AIEC) to host cells, we considered them to be extracellular. Host proteins were also selected based on their localization information.

For compiling the extracellular list, selection was done by retaining only proteins located in "plasma membrane" and "extracellular region" according to the databases LocDB (Rastogi and Rost, 2011) and Human Protein Atlas (HPA) (Thul and Lindskog, 2018). Proteins corresponding to all the remaining localization terms were considered as intracellular.

Gene expression in CD patients

The set of differentially expressed genes (DE-Gs) between CD and non-IBD subjects were retrieved from Lloyd-Price et al. (2019). Since biopsies were taken from two different sites (rectum and ileum), we obtained independent lists of differentially expressed genes for each of the two sites.

Inferring effects of microbe-host predictions on host gene expression

To infer the most probable signaling chains which capture the functional effects of the bacterial proteins on host processes, we used the network diffusion algorithm TieDie (Paull et al., 2013). TieDie not only discounts for topological biases in networks such as hubs but also mines out logically consistent paths between a given set of start and stop nodes. Host proteins which are predicted to be bound by the bacterial proteins were used as the start nodes. As for the stop nodes, the transcription factors (TFs) of the differentially expressed genes in CD patients were used. TFs of the DE-Gs were retrieved from the set of directed and signed transcriptional regulatory interactions from DoRothEA (Garcia-Alonso et al., 2019). For each of the TFs in the set of stop nodes, weights were calculated using the following formula (Equation 1). The equation considers both the fold change of the DEG as well as the sign (i.e stimulatory or inhibitory) of the relationship between the TF and the DEG. The set of directed and signed protein-protein interactions from OmniPath(Türei et al., 2016) were used as a source of interactions for inferring the signaling chains between the start and the stop nodes.

$$weight_{TF} = \frac{1}{\#TG_{TF}} \sum_{1}^{TG} FG_{TG} * signTFTG \qquad \text{(Equation 1)}$$

Retrieval of biologically plausible chains

Given that there exist upper limits for the diameter of real-world biological networks, we set out path length limitations in order to capture the most relevant signaling chains inferring from the network diffusion methodology described above. In the case of signaling chains mediated by the set of extracellular bacterial proteins, we set a path length of four (i.e., the bacterial protein followed by the host receptor, an intermediary host protein, the TF and the DEG). As for the set of intracellular bacterial proteins (all from AIEC), we set a path length of three (i.e., the bacterial protein followed by the host receptor, the TF and the DEG). Paths of the desired lengths (as described above) were extracted using the command line version of PathLinker (Ritz et al., 2016).

Network diffusion and chain extractions were performed separately for every bacterial species and the two sets of DE-Gs corresponding to both the ileal and rectal biopsies. Only chains not corresponding to interactions between bacterial and host proteins mediated by structural complex formation events were prioritized in order to retain the chains specific to post-translational modifications elicited by the bacterial proteins on the host proteins. Furthermore, chains corresponding to post-translational modifications elicited by bacterial histidine kinases (HKs) were discarded since the phosphorylation domains of HKs are predominantly embedded in the cytoplasmic-membrane boundary of bacterial cells (thus rendering the HKs incapable of inducing phosphorylation events on host proteins).

Functional enrichment analysis

The R packages ReactomePA (Yu and He, 2016) and clusterProfiler (Yu et al., 2012) were used to perform the enrichment analysis. An adjusted P-value cut-off of ≤ 0.1 was used to infer the statistically significant enrichment terms. For the enrichment analysis of the signaling networks, all the nodes in the directed PPI network were set as the background. For the corresponding analysis of the DE-Gs in the ileal signaling network, the entire set of DE-Gs were used as the background. For the identification of site-specific enrichment terms, the enrichment was performed with different gene sets separated by site and gene categories. All human genes (receptors, intermediary proteins, transcription factors (TFs) and DE-Gs) were considered for the site-specific enrichment analysis.

Validation cohort

A cohort of paired 16S read-outs from stool samples (n = 20) and gene expression data (n = 20) from ileal biopsies in CD patients, cross-sectionally collected at the University Hospitals Leuven was available for independent validation. The ethics committee of the University Hospitals Leuven approved the study (IRB approvals, B322201213950/S53684 and B322201627472/S57662). All individuals gave written informed consent.

Microbial load measurement by flow cytometry

Cell counting for all samples, was performed in duplicate. Briefly, 0.2 g frozen (−80°C) aliquots were dissolved in physiological solution to a total volume of 100 mL (8.5 g/L NaCl; VWR International, Germany). Subsequently, the slurry was diluted 1,000 times. Samples were filtered using a sterile syringe filter (pore size of 5 μμm; Sartorius Stedim Biotech GmbH, Germany). Next, 1 mL of the microbial cell suspension obtained was stained with 1 μL SYBR Green I (1:100 dilution in DMSO; shaded 15 min incubation at 37°C; 10,000 concentrate, Thermo Fisher Scientific, Massachusetts, USA). The flow cytometry analysis was performed using a C6 Accuri flow cytometer (BD Biosciences, New Jersey, USA) based on Prest et al. (Doherty et al., 2018) Fluorescence events were monitored using the FL1 533/30 nm and FL3 >670 nm optical detectors. In addition, also forward and sideward-scattered light was collected. The BD Accuri CFlow software was used to gate and separate the microbial fluorescence events on the FL1/FL3 density plot from the faecal sample background. A threshold value of 2000 was applied on the FL1 channel. The gated fluorescence events were evaluated on the forward/sideward density plot, as to exclude remaining background events. Instrument and gating settings were kept identical for all samples (fixed staining/gating strategy (Doherty et al., 2018)). Based on the exact weight of the aliquots analysed, cell counts were converted to microbial loads per gram of faecal material. Faecal Moisture content was determined as the percentage of mass loss after lyophilization from 0.2 g frozen aliquots of non-homogenized faecal material (−80°C).

DNA extraction and sequencing data pre-processing

Faecal DNA extraction and microbiota profiling was performed as described previously (Gevers et al., 2014). Briefly, DNA was extracted from faecal material using the MoBio PowerMicrobiome RNA isolation kit with the addition of 10 min incubation at 90°C after the initial vortex step. The V4 region of the 16S rRNA gene was amplified with primer pair 515F/806R (Caporaso et al., 2011). Sequencing was performed on the Illumina MiSeq platform (San Diego, California, USA), to generate paired-end reads of 250 bases in length in each direction. After de-multiplexing, fastq sequences were merged using FLASH (Magoč and Salzberg, 2011) software with default parameters. Successfully combined reads were filtered based on quality (>90% of nucleotides with quality score of 30 or higher for every read) using Fastx tool kit (http://hannonlab.cshl.edu/fastx_toolkit/). Chimeras were removed with UCHIME (Edgar et al., 2011). Faecal samples were processed altering the protocol above to dual-index barcoding. Pre-processing was performed using the DADA2 (Callahan et al., 2016) pipeline v1.6.0.

Relative microbiome profiling (RMP)

For the relative microbiome matrix, each sample was downsized to 10,000 reads by random selection of reads. Samples with less than 10,000 reads were excluded (two samples). The taxonomy of reads was assigned using RDP classifier 2.12 (Wang et al., 2007).

Quantitative microbiome profiling (QMP)

The quantitative microbiome profiling matrix was built as described by Vandeputte and colleagues (Vandeputte et al., 2017). In short, samples were downsized to even sampling depth, defined as the ratio between sampling size (16S rRNA gene copy number corrected sequencing depth) and microbial load (average total cell count per gram of frozen faecal material). 16S rRNA gene copy numbers were retrieved from the ribosomal RNA operon copy numberdatabaserrnDB(Stoddard et al., 2015). The copy number corrected sequencing depth of each sample was rarefied to the level necessary to equate the minimum observed sampling depth in the cohort. Samples with resulting rarefied read counts <150 were excluded from QMP analyses. Rarefied genus abundances were converted into numbers of cells per gram.

Relative microbiome sequence variant profiling

Characterization of the microbiota profiles below genus-level was performed using the DADA2 (Naftali et al., 2016) pipeline sequence variants, with taxonomy assignment by RDP classifier v 2.12 (Wang et al., 2007). A phylogenetic tree for the sequence variants was reconstructed using the dada2 sequence variant

sequences, by maximum likelihood reconstruction (GTR model) using PhyML (Guindon et al., 2010). Sequence variants were grouped into species clusters by collapsing sequence variants with less than 0.005 branch length distance in the sequence variants tree. Competition between species clusters with increasing inflammation levels was evaluated by assessing the correlation between calprotectin levels and species clusters dominance, i.e. their relative proportion of the total genus abundance per sample, excluding samples for which the genus total abundance sum was zero.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed in R using packages *vegan* (Oksanen, 2021), *phyloseq* (McMurdie and Holmes, 2013), *FSA* (Derek, 2021), *coin* (Hothorn, 2021), *DirichletMultinomial* (Morgan, 2021), and *fitdistrplus* (Delignette-Muller, 2021). All statistical tests used were two-sided. All p values were corrected for multiple testing using the Benjamini-Hochberg method (reported as FDR) when multiple tests were performed on lists ($n > 1$) of features (e.g. taxa-metadata or taxa-taxa associations), also when performing multiple pairwise group ($n > 2$) comparisons (e.g. Kruskal-Wallis test with post-hoc Dunn test). Cramer's phi ($\sqrt{\chi^2}/N$) was used as measure of effect size for chi square tests (likelihood ratio test) and $r = Z/\sqrt{N}$ for the Wilcoxon signed rank test.

Faecal microbiome derived features and visualization

Observed genus richness was calculated on the RMP matrix (downsized to 10,000 reads as described above) using *phyloseq* (McMurdie and Holmes, 2013). Enterotyping (or community typing) based on the Dirichlet Multinomial Mixtures (DMM) approach was performed in R as described by Holmes et al. (Holmes et al., 2012). Enterotyping was performed on a combined genus-abundance matrix including IBD and healthy controls samples compiled with the samples originating from the Flemish Gut Flora Project. Microbiome inter-individual variation was visualized by principal coordinates analysis (PCoA) using Bray-Curtis dissimilarity on the genus-level abundance matrix.

Microbiome and physiological features associations

Taxa unclassified at genus level or present in less than 20% of samples were excluded from the statistical analyses. Pearson or spearman correlations were used respectively for linear or rank-order correlations between continuous variables. Mann–Whitney U was used to test median differences of continuous variables between two different groups. For more than two groups, Kruskal-Wallis test with post-hoc Dunn test were used. Statistical differences in the proportions of categorical variables (enterotypes) between patient groups were evaluated using pairwise Chi-square tests.

Transcriptomic analysis

Only CD patients ($n = 20$) with a Bacteroides 2 enterotype were included for the differential gene expression analysis from their corresponding ileal biopsies (Table S13). A non-IBD group was used as the control ($n = 15$). RNA extraction and sequencing was performed as reported in (Verstockt et al., 2019). Raw RNA-sequencing data were aligned to the reference genome using Hisat2 version 2.1.0 (Kim et al., 2015) and absolute counts generated using HTSeq (Anders et al., 2015). Counts were then normalized for library size using the DESeq2 package (Love et al., 2014), and only protein-coding genes (Ensemble hg 19 reference build) having an average of at least 10 normalised read counts were considered for further analysis ($n = 14,263$). Differential expression analysis with multiple testing correction (Benjamini-Hochberg method) was then performed, and correction for age and gender was taken into account. Genes with a log2 fold change of ≥ 1.5 and ≤ -1.5 and an adjusted P-value of ≤ 0.05 were considered to be differentially expressed. The raw datasets have been deposited at ArrayExpress (accession number: E-MTAB-10395).

Drug and disease relevance of the networks

Proteins ($\# = 5484$) known to be associated with CD based were retrieved from the OpenTargets (Carvalho-Silva et al., 2019) database (Table S12). Only those proteins ($\# = 720$) with at least two different sources of evidence were considered for further analysis. To identify the possible drugs or perturbations which can mimic or reverse the signatures of the DE-Gs (in the CD patients) in the microbiota-induced ileal and rectal networks, the DEG signatures were matched against the Library of Integrated Network-based Cellular Signatures (LINCS) L1000 dataset (Duan et al., 2016). The LINCS L1000 dataset contains more than a million gene expression profiles of human cell lines treated with chemical perturbations or drugs.