

RESEARCH

Open Access



A systematic review and functional bioinformatics analysis of genes associated with Crohn's disease identify more than 120 related genes

Debora Garza-Hernandez¹, Maricruz Sepulveda-Villegas¹, Jose Garcia-Pelaez^{2,3}, Raul Aguirre-Gamboa⁴, Peter L. Lakatos⁵, Karol Estrada⁶, Manuel Martinez-Vazquez⁷ and Victor Trevino^{1,8*}

Abstract

Background: Crohn's disease is one of the two categories of inflammatory bowel diseases that affect the gastrointestinal tract. The heritability estimate has been reported to be 0.75. Several genes linked to Crohn's disease risk have been identified using a plethora of strategies such as linkage-based studies, candidate gene association studies, and lately through genome-wide association studies (GWAS). Nevertheless, to our knowledge, a compendium of all the genes that have been associated with CD is lacking.

Methods: We conducted functional analyses of a gene set generated from a systematic review where genes potentially related to CD found in the literature were analyzed and classified depending on the genetic evidence reported and putative biological function. For this, we retrieved and analyzed 2496 abstracts comprising 1067 human genes plus 22 publications regarding 133 genes from GWAS Catalog. Then, each gene was curated and categorized according to the type of evidence associated with Crohn's disease.

Results: We identified 126 genes associated with Crohn's disease risk by specific experiments. Additionally, 71 genes were recognized associated through GWAS alone, 18 to treatment response, 41 to disease complications, and 81 to related diseases. Bioinformatic analysis of the 126 genes supports their importance in Crohn's disease and highlights genes associated with specific aspects such as symptoms, drugs, and comorbidities. Importantly, most genes were not included in commercial genetic panels suggesting that Crohn's disease is genetically underdiagnosed.

Conclusions: We identified a total of 126 genes from PubMed and 71 from GWAS that showed evidence of association to diagnosis, 18 to treatment response, and 41 to disease complications in Crohn's disease. This prioritized gene catalog can be explored at <http://victortrevino.bioinformatics.mx/CrohnDisease>.

Keywords: Crohn's disease, Inflammatory bowel disease, Genes, Functional variants, Mutations

Background

Inflammatory bowel diseases (IBD) comprise Crohn's disease (CD) and ulcerative colitis (UC), which are inflammatory diseases of the gastrointestinal tract with an unknown etiology [1]. Common symptoms of CD include abdominal pain, fever, diarrhea, and bleeding, depending on disease severity [2]. Disease complications

*Correspondence: vtrevino@tec.mx

¹ Tecnológico de Monterrey, Escuela de Medicina, Cátedra de Bioinformática, Av. Morones Prieto No. 3000, Colonia Los Doctores, 64710 Monterrey, Nuevo León, Mexico

Full list of author information is available at the end of the article



can lead to bowel disability and sometimes to surgery [3]. CD is more frequent among industrialized nations such as North America, with a reported incidence of 6.3 to 23.8 per 100,000, and Western Europe, with 1.9 to 10.5 per 100,000 people [4, 5].

Therefore, in addition to common risk factors for CD, the contribution of genetic factors in CD has been considered highly relevant. This contribution is based on the fact that family history can influence the presence of the disease, with a higher risk for siblings with a relative risk of 13 to 36 times [6]. In fact, heritability estimates for Crohn's disease from pooled twin studies have been reported to be 0.75 [7].

As with many other complex traits [8], several CD related-genes have been identified through the use of linkage-based studies, candidate gene association studies (i.e., transmission disequilibrium tests), and high coverage technologies such as DNA arrays and next-generation sequencing (NGS) [9, 10]. Among well-known risk genes for CD are *NOD2*, *IL23R*, and *ATG16L1* [11], which are involved in inflammation and the immune system's response [11, 12].

In addition to candidate genes association studies, the implementation of high coverage technologies, such as NGS, has improved the molecular diagnostic yield of complex diseases such as CD. These strategies typically make use of phenotype-specific panels containing genes that are known to confer susceptibility for a complex disease [13, 14]. Specifically, for CD there have been attempts to test for genetic susceptibility for treatment response and prognosis in CD patients [15, 16]. Genome-wide association studies (GWAS) have also made large contributions identifying more than 130 genes [17–19]. From these, genome-wide polygenic risk scores (PRS) aim to identify individuals at significantly increased risk. For CD, PRS from over 200 loci, yields an estimate of 8% of variance explained and an AUC around 0.7 [20].

A comprehensive collection of genes for CD is lacking, which complicates further functional analyses and overall understanding. Different aspects of CD have been reviewed, including inflammatory drugs and risk of exacerbation [21], pouch incidence [22], prognostic factors [23, 24], and biomarkers for surgery outcomes [25]. Nevertheless, to our knowledge, there is a lack of functional analyses and systematic reviews analyzing all known genes or variants associated with CD susceptibility. We conducted this compilation by first classifying each gene based on the genetic evidence reported and then functionally analyzing those genes. We hope this collection of genes and functional analysis might help for further understanding of the disease etiology.

Starting from a Pubmed query, we systematically curated 2496 abstracts following recommended

methodologies to identify and functionally classify genes associated with CD. To further support our findings, we provided functional analyses of the identified genes. We show that although most of the research in CD revolves around a group of well-known genes, our systematic curation review identified 126 genes with a sufficient level of associative evidence.

Methods

Based on our previous work [26], we collected abstracts related to genetic variations in CD from the PubMed repository. The following review process adheres to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) 2020 guidelines [27]. The abstracts were manually revised, curated, and annotated for each identified gene by using the PubTerm web tool [28]. Each gene was designated to a specific category based on the reported genetic alteration or evidence related to CD. The details are described in the next sections.

Abstract collection

Only original research papers published in English were considered. The search strategy comprised three basic terms: (1) Crohn's disease, (2) genetic variations, and (3) focus on humans. Thus, the following query was used: *crohn*[TIAB] AND (mutation*[TIAB] OR polymorphism*[TIAB] OR variant*[TIAB]) NOT review[Publication Type] NOT mouse[TIAB] NOT mice[TIAB]*. The query was performed during 2019 and updated in January 2020. We used PubTerm to curate and annotate abstracts per gene, previously used in pulmonary arterial hypertension and vitamin D levels [26, 29]. Additionally, we reviewed GWAS publications as described below.

Definition of gene categories

We defined categories to annotate the genes identified based on the genetic evidence related to CD ordered by importance as: (i) *Experimental evidence of a variant*, when experimental evidence of specific sequence variants is shown for CD; (ii) *GWAS evidence within gene*, if sequence variants or single nucleotide polymorphisms (SNP) were found within a gene region in GWAS; (iii) *Genetic evidence in treatment response*, when experimental evidence of sequence variants were associated with response to CD treatments; (iv) *Genetic evidence in related complications*, if experimental evidence of sequence variants were associated with CD complications; (v) *Other genetic alterations*, when no specific sequence variant information was provided (e.g. haplotypes, SNP at intergenic regions, uncertain locus); (vi) *Genetic evidence in a related disease*, when experimental

evidence of sequence variants is shown for other related diseases rather than specifically to CD; (vii) *Related but not variant reported*, if no genetic evidence is shown but there is a biological relationship mentioned between the gene and the disease (e.g. gene expression changes); (viii) *Negative evidence*, if the gene is properly annotated but the conclusion of the research was a not causal relationship; (ix) *Unrelated*, if the gene is correctly annotated but there is no mention of the causal association of the gene with the disease; (x) *Annotation error*, if the gene is not related to CD due to nomenclature errors, inaccurate disease, and other diverse errors. *SNX20* gene was added manually, as our search identified a paper with evidence of its association with CD, but the *SNX20* gene symbol was not correctly identified by the tools used.

Curation and categorization per gene

After the abstracts were retrieved from PubMed into the PubTerm tool [28], we filtered for only human genes, and each gene was subsequently reviewed. All the abstracts organized per gene were carefully read and analyzed until enough evidence was convincing to assign the gene to a specific category, or all abstracts were carefully read. If two categories apply, the category with more relevant genetic information was used. The full text was reviewed when necessary, commonly when a sequence variant was not clear, uncertain, or in negative cases. The critical sentence in the abstract and the PubMed ID was added to the PubTerm notes in every gene analyzed to support the decision made, which is available electronically within PubTerm, as shown below. Most genes were reviewed by two authors. All results can be obtained from Supplementary Table 1 and PubTerm (<http://victortrevino.bioinformatics.mx/pubterm>) using the user “vtrevino@tec.mx” and project “Crohn_s Disease”. In addition, to facilitate rapid revision, we provide a summary list at <http://victortrevino.bioinformatics.mx/CrohnDisease>.

GWAS variants revision

A search for GWAS studies was performed at GWAS Catalog [30] in order to retrieve variants that were not mentioned and indexed directly into PubMed abstracts and full texts. For this, only publications with reported associations specifically for CD were used. This was done using the search term Crohn’s disease trait with EFO_0000384. Also, a comparative search was performed at the Open Targets platform [31], which integrates public domain data to enable finer target identification and prioritization for a given disease. For comparative purposes, only the genetic associations data type was used. This data comes from a Linkage-disequilibrium expansion and fine mapping of GWAS curated associations. Thus, it aims to identify the most likely causal variant

linked to the GWAS detected variant. If a gene found in GWAS and PubMed, the higher categorical evidence was kept.

Identification and annotation of variants

To further support our findings, the variants were also reviewed in ClinVar [32]. For genes not reported in ClinVar, a manual annotation approach was performed by using the information from the original publication. The list of variants and their respective transcript or “rs ID numbers” (for SNPs) is presented as supplementary information.

In-silico functional analysis of genes

For the CD confirmed genes ($n=126$), a functional analysis was carried out using DAVID [33]. This tool performs an over-representation test to determine, from an input set of genes, if the number of genes appearing in a biological pathway or biological term is not random. In such a case, the gene set is said to be tightly associated with the term. We used a hierarchical clustering approach to group the biological terms obtained from DAVID for comparison and summarization purposes. The functional analysis performed consisted of Gene Ontology terms (including *cellular component*, *biological processes*, and *molecular functions* [34, 35], KEGG pathways, and related diseases from the genetic association database (GAD) [36]. The criteria for clustering terms consisted of selecting those terms statistically significant (after Bonferroni correction for $p < 0.05$) and that involved a large number of genes (≥ 12 for diseases, ≥ 10 for GO and KEGG). Manual merging was also performed to group similar concepts and hence facilitated the interpretation. Because many highly related terms were observed, the significant terms were grouped by similarity using hierarchical clustering separately for GO and KEGG and by groups of similar disorders or diseases. Groups were generated by averaging the presence of the gene among the diseases/terms merged in the group.

Additionally, we used the Gene Network v2.0 tool [37] to identify Human Phenotype Ontology (HPO) clusters [38]. We selected the most distinguishable clusters based on co-regulation scores across public RNA-Seq samples and ran a phenotype analysis for each selection. Only terms with a significant enrichment (Bonferroni $p < 0.05$) were selected.

For the differential expression, the 126 genes associated with CD were analyzed within the recently published Gene Expression Omnibus dataset [39] GSE111889, which shows recent data from UC and CD compared to normal ileum and colon. Differential expression was performed separately for CD, UC, and tissue. A linear model regression with sex correction

was fitted for each analysis. Differentially expressed genes (DE) were selected after a false discovery rate correction <0.05 [40].

Gene-drug interactions were analyzed using the drug-gene interaction database (DGIdb) [41] using the default parameters.

Benchmark on genes present on commercial panels and identified by GWAS

To show the relevance of the extracted data and possible applications, a comparison was performed between the genes extracted from our curation process and the Genetic Test Registry (GTR) [42]. For GTR, diagnosis panels for CD and related diseases were tested. The clinical panels used for comparison were searched for the IBD1: Crohn Disease, the criteria for selecting the tests consisted of selecting only specific diagnostic tests for CD or inflammatory-related diseases. A general web search was also performed with a search strategy comprising the queries “Crohn’s disease testing panels”, “Crohn’s disease genetic diagnosis panels”, and “Crohn’s disease diagnosis commercial panels”.

Results

Classification and identification of CD genes

The PRISMA flow chart for the selection of studies and genes is shown in Fig. 1. The PRISMA checklist is provided as supplementary information (File S1). The PubMed search imported into PubTerm identified 2496 articles, which referred to 1172 genes. The genes were reduced to 1055 after filtering for human genes (Fig. 1, Table S1). Then, each gene was carefully curated, annotated, and categorized as described in methods. The curation revealed that 400 genes were somehow potentially related to CD while 655 genes were not related due to annotations errors, the gene mention was casual, no association was found, or no evidence of variation was shown (mainly in subsequent gene expression changes). From the 400 genes, 81 were finally categorized as associated with a *related disease* such as IBD in general, UC, familial diarrhea syndrome, colorectal cancer, or chronic lymphocytic leukemia. Ninety-three genes were classified to *other genetic alterations* because its gene was uncertain, which included genes identified through haplotypes or intergenic regions in GWAS. Thus 226 genes were confirmed as associated with CD from this curation (Fig. 1).

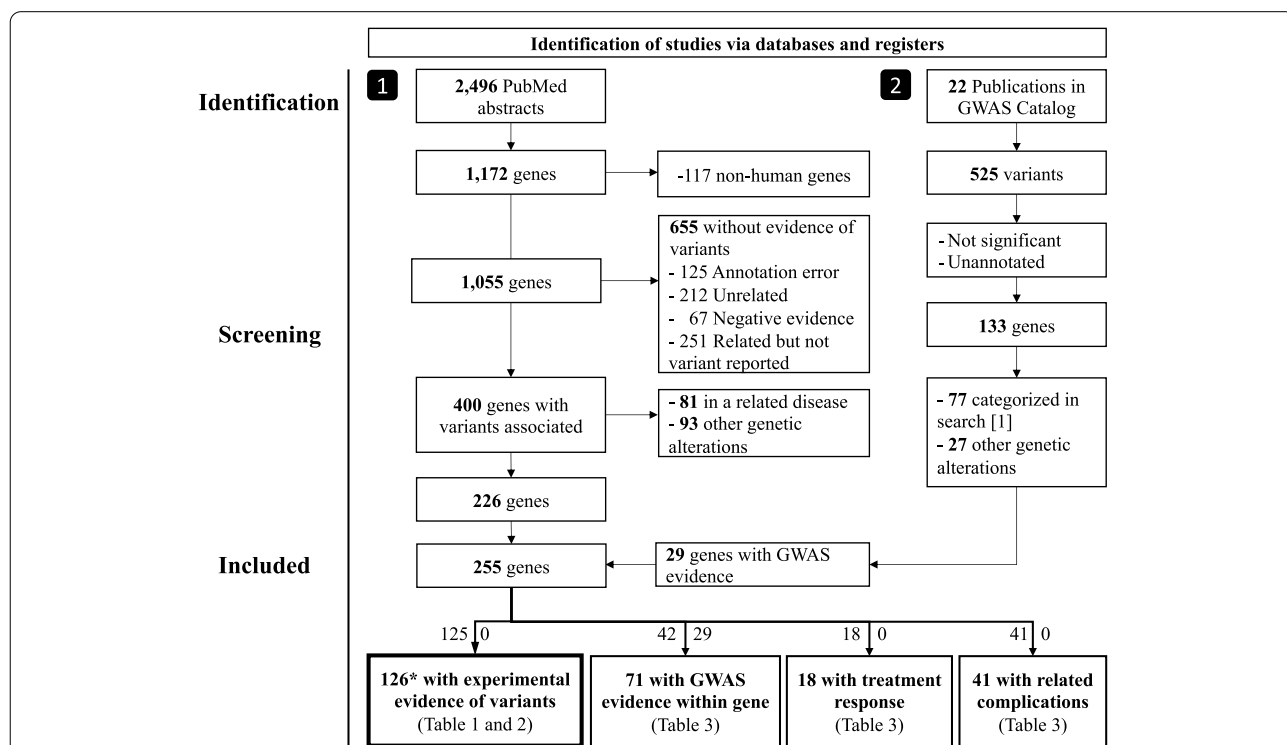


Fig. 1 Summary of categorized genes for Crohn’s disease. The numbers at left in arrows at the bottom represent the genes from search [1], while the numbers at right correspond to search [2]. *SNX20 was added manually

We also used GWAS Catalog [30] as a source of gene information. From 22 publications for CD risk, we obtained the 525 comprising variants. Variants were further filtered by removing those whose tagged gene was not reported, were not significant, or the gene or variant was duplicated, leaving only 133 genes (Table S2). From these, 77 were already categorized in the PubMed curation described above. Thus 56 genes were added to our list of genes, 27 intergenic variants were assigned to *other genetic alterations*, and 29 to *GWAS evidence within gene*. A list of the considered PubMed abstracts is shown in Supplementary Information (files S2 and S3). In summary, we identified 256 genes associated with diverse aspects of CD (Fig. 1).

A total of 126 genes were found to have *experimental evidence of variants* in CD. The top 26 genes of this category mentioned in more than 15 abstracts are shown in Table 1, while the genes with less than 15 abstracts are summarized in Table 2 and detailed in the Supplementary Information (Table S4). The topmost frequent genes for this category are well-known for their association with CD [11, 12], such as *NOD2*, *TNF*, *IL23R*, *ATG16L1*, *TLR4*, *IL10*, *SLC22A4*, *SLC22A5*, and *IRGM* (Table 1).

Besides the above 126 genes (Tables 1 and 2), we also found 71 genes associated with CD that were categorized as *GWAS evidence within gene* where an SNP is located within genomic coordinates, either an intron or exon (Table 3). Additionally, 18 genes were found to be specifically associated with *treatment response* in CD and 41 related to *disease complications* (Table 3).

Location of the functional variants

Of the 126 genes corresponding to the categories of *experimental evidence of variants* plus 71 genes with *GWAS evidence within gene*, only 17 genes (<12%) were found to be annotated in ClinVar [32]. In this context, to support our systematic categorization, a supplementary file is provided with the information referring to the location of the variants that were not found in ClinVar (Tables S4 and S5). This information was revised and obtained either by the original paper or by the information related to the SNP reported at dbSNP [43].

In-silico functional analysis

To provide an overview of the 126 genes with experimental evidence to CD, a functional bioinformatics analysis was performed [33, 44]. For this, we assessed whether the genes prioritized in our study are indeed statistical and biologically relevant. We performed gene set enrichment analysis in multiple databases containing different biological terms, including pathways (KEGG), diseases (GAD), and gene ontology (GO) terms. Detailed results of all enriched gene sets are

Table 1 The subset of top genes with experimental variants associated with CD (abstracts > 15)

Gene	Abstracts	Panels	GWAS catalog studies	ClinVar
<i>NOD2</i>	751	16	8	Yes
<i>TNF</i>	156	0	0	Yes
<i>IL23R</i>	148	1	6	Yes
<i>ATG16L1</i>	128	1	5	Yes
<i>TLR4</i>	88	0	0	No
<i>IL10</i>	86	2	1	Yes
<i>SLC22A4</i>	73	0	0	Yes
<i>IRGM</i>	60	1	3	Yes
<i>SLC22A5</i>	50	0	0	Yes
<i>TNFSF15</i>	45	0	4	No
<i>NOD1</i>	41	0	0	No
<i>IL6</i>	41	4	0	Yes
<i>IL1B</i>	40	0	0	No
<i>STAT3</i>	37	1	2	No
<i>NFKB1</i>	36	0	0	No
<i>DLG5</i>	35	0	0	No
<i>IL12B</i>	34	0	0	No
<i>ABCB1</i>	33	1	0	Yes
<i>KRAS</i>	33	0	0	No
<i>PTPN22</i>	31	0	0	No
<i>IL1RN</i>	29	0	0	No
<i>IL23A</i>	28	0	0	No
<i>CD14</i>	28	0	0	No
<i>PTPN2</i>	24	0	0	No
<i>IL10RA</i>	20	2	0	Yes
<i>NLRP3</i>	19	0	0	No
<i>MEFV</i>	20	1	0	No
<i>IL4</i>	18	0	0	No
<i>NKX2-3</i>	17	0	0	No
<i>ICAM1</i>	16	0	1	No
<i>IFNG</i>	16	0	0	No
<i>TLR9</i>	16	0	0	No

present in the Supplementary Information. Because the number of significant terms was high, repetitive, and difficult to interpret, we grouped the terms by biological and genetic similarity (see [Methods](#)).

Regarding diseases, as expected, the most similar term to CD is IBD and UC, validating our strategy (Fig. 2). We observed a dense group of genes and diseases where CD is located close to other groups of autoimmune and inflammatory diseases, certain types of cancer, hypersensitivity disorders caused by allergies and intolerance, infections by virus and bacteria, pregnancy complications, and metabolic complications.

Gene sets from GO are divided into cellular components, biological processes, and molecular functions.

Table 2 Genes with experimental variants associated with CD mentioned by less than 15 abstracts. Details are provided in Table S1. * denotes manual addition

ACE	CFTR	EPX	IL10RB	MST1	PTGS2	TLR6
AGER	CLEC2D	ERAP2	IL16	MTRR	REEP6	TNFAIP3
AGT	CREM	FAS	IL18	MUC2	SFTPD	TNFRSF1A
APOE	CSF1R	FUT2	IL27	MUC3A	SLC11A1	TNFRSF1B
ATG16L2	CSF2RB	GC	IL4R	MYO9B	SLC15A1	TNFSF8
BDKRB1	CX3CR1	GSTT1	IRF1	NAT2	SLC22A1	TRAIP
BPI	CXCL16	HLA-DQA2	IRF5	NCF2	SLC39A8	UCP2
BTNL2	CXCR4	HLA-G	JAK2	NCF4	SLCO3A1	ULK1
CALCOCO2	CYP2A6	HNFA4	KCNN4	NFKBIA	SMAD3	XIAP
CARD9	DEFB1	HNRNPD	KRT8	NOS2	SNX20*	ZNF365
CCR2	DLG1	HSPA1L	MAG3	ORMDL3	TCN2	
CCR5	DMBT1	HSPA4	MAP3K8	POU5F1	TIMP1	
CD24	DNAH12	IFNA10	MIF	PPARG	TLR1	
CD40LG	DUOX2	IFNA4	MLN	PTEN	TLR5	

Within biological processes terms, we observed significant enrichment in processes related to *response to bacteria*, *positive regulation of nitric oxide (NO)*, *ERK*, and *NFκβ*, *apoptosis*, *cell proliferation*, *response to lipopolysaccharide*, *inflammatory* and *immune response*. For molecular function, *cytokine activity*, *interleukine-1 receptor binding*, *receptor activity*, and *protein homodimerization activity* were significantly enriched by the CD-risk genes analyzed. For cellular components, only *membrane*, *plasma membrane*, and *extracellular region* were significant. Overall the significantly enriched GO terms point to known CD terms such as the immune response, cytokine activity, and signaling receptors as the primary source of functional causes (in terms as *Immune Response*, *Infection*, *Interleukin-1 receptor binding*, *Cytokine and defense response*, *Cytokine activity*, *Autoimmune Disease*).

Additionally, among the enriched pathways identified were *JAK-STAT signaling pathway*, *cytokine receptor interaction*, *NOD-like receptor signaling pathway*, *NF-kappa*, *TNF signaling*, *Toll-like receptor (TLR) signaling*, *T cell receptor signaling*, and *Osteoclast differentiation*. These signaling pathways converge on the activation of NF-κB, a protein complex that controls the transcription of DNA cytokine production and cell survival [45]. In addition, the KEGG comorbidities identified are *infectious diseases caused by bacteria, protozoa and virus*, and *IBD*, which are reliable associations due to the relationship between microbes and CD [46]. The mapping is, therefore, an excellent guide to connect genes and important biological aspects of CD (Fig. 2).

Among the genes present in the enriched biological terms, we noted two distinct groups depending on the frequency of their presence in the gene sets and

their number of abstracts found by PubTerm, designated as *common* and *sporadic* (Fig. 2). Briefly, the 34 *common* genes are highly related to diseases and biological terms and well-studied. In comparison, the 92 *sporadic* genes are associated with particular diseases or biological terms and not as studied in CD as the *common* genes. Among the *common group*, the most shared genes across concepts are *NOD2*, *TNF*, *ICAM1*, *NFKBIA*, *NFKB1*, *TNFRSF1A*, *CD14*, *ACE*, *TLR4*, and *TLR9*, which are involved in both TNF and of NF-κβ signaling pathways [47–49]. Also, *IFNG*, *IL1B*, *IL6*, *IL23R*, *IL10*, *IL4*, *IL12B*, *IL1RN*, *IL18*, and *IL4R*, which are all well-known cytokines or related genes, and contribute to the inflammatory response and cytokine interaction process [50–52]. The *sporadic* group comprised 92 genes that were much less frequent among enriched terms, where *ATG16L1*, *SLC22A4*, *IRGM*, *SLC22A5*, *TNFSF15*, *NOD1*, *PTPN2*, *PTPN22*, and *DLC5* being more frequently mentioned.

DNAH12, *ERAP2*, *FUT2*, *ORMDL3* and, *TRAIP* were more specifically enriched in the CD disease term and less common for the remaining terms.

Once we noted two clear sets of *common* and *sporadic* genes that were associated with specific terms, we considered whether the genes might also be grouped by other phenotypes that could explain CD symptoms. Thus, we used the Gene Network tool, which clusters genes with similar Human Phenotype Ontology. Five sub-networks were identified from the 126 genes categorized as *experimental evidence*. We then merged three highly interconnected sub-networks and subsequently analyzed the three resultant modules (73, 26, and 26 genes respectively for Module 1 in green, Module 2 in purple, and Module 3 in blue, as shown in Fig. 3). Next, for

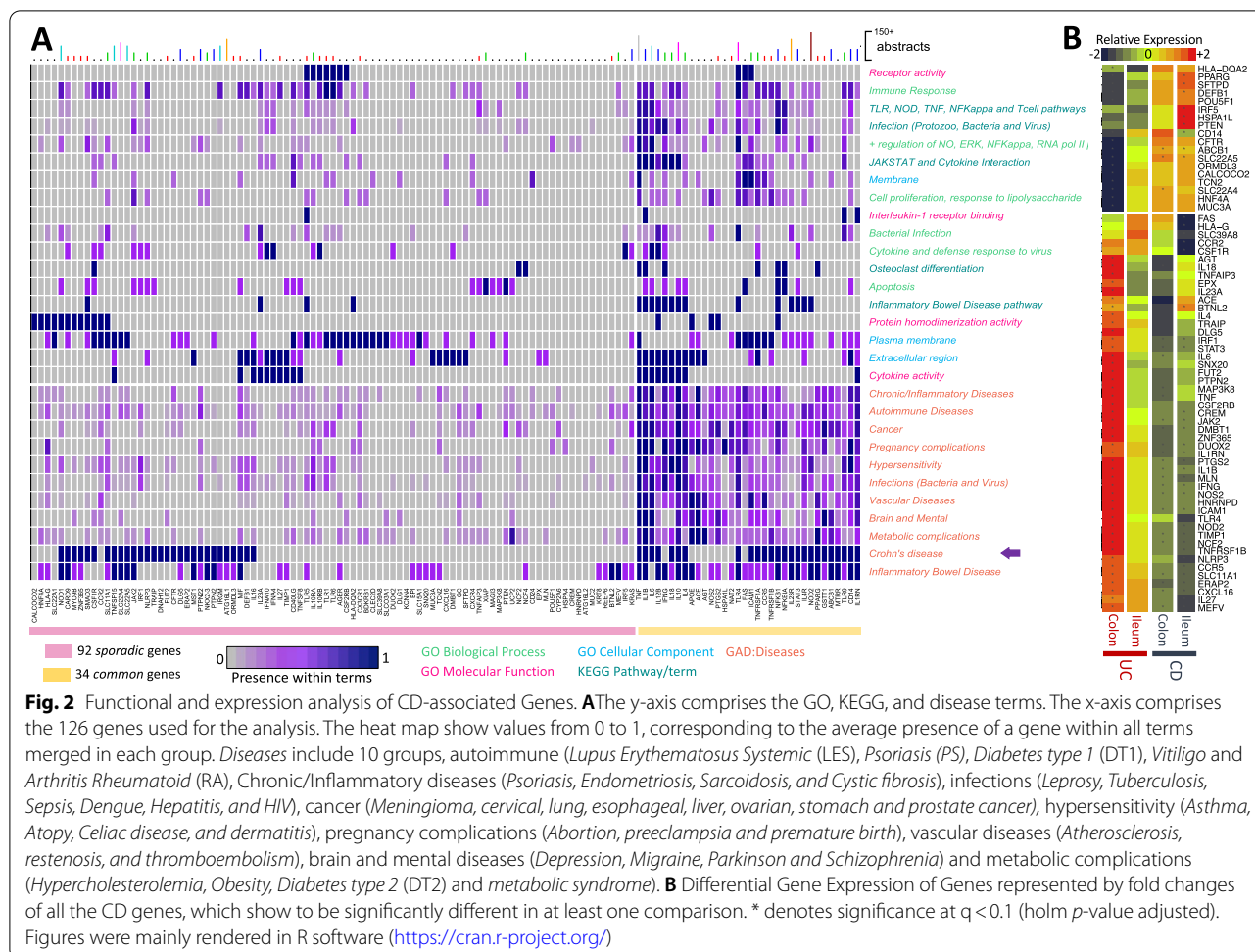
Table 3 Other genes associated with CD for diverse categories. * Retrieved from GWASCatalog. + Retrieved from a panel

Category	Feature	Genes (Abstracts)
GWAS evidence within the gene	SNP within gene associated to CD from GWAS	<i>LRRK2</i> (19), <i>STAT4</i> (14), <i>IFNGR2</i> (3)*, <i>CYLD</i> (2), <i>CLEC16A</i> (2), <i>LY75</i> (2), <i>TLR8</i> (2), <i>ZMIZ1</i> (2), <i>CXCL12</i> (2), <i>ZAP70</i> (2)+, <i>TLR10</i> (2), <i>PER3</i> (2), <i>MTMR3</i> (2), <i>MAG1</i> (2), <i>CD40</i> (2)*, <i>TAB1</i> (1), <i>CDYL2</i> (1), <i>ELF1</i> (1)*, <i>NLRP11</i> (1), <i>IL2RB</i> (1), <i>MAP3K1</i> (1), <i>TIMMDC1</i> (1), <i>PDE2A</i> (1), <i>PRKCO</i> (1), <i>SLC22A23</i> (1), <i>TLE1</i> (1), <i>TRPM2</i> (1), <i>MORC4</i> (1), <i>CYP4F2</i> (1), <i>CLCA2</i> (1), <i>SLC23A1</i> (1), <i>GCKR</i> (2)*, <i>IL18RAP</i> (4)*, <i>BRD2</i> (2)*, <i>INAVA</i> (1)*, <i>IL2RA</i> (4)*+, <i>SP140</i> (1)*, <i>ITLN1</i> (1)*, <i>BACH2</i> (4)*, <i>GPR65</i> (2)*, <i>IL1RL1</i> (4)*, <i>PUS10</i> (1)*, <i>ANKRD55</i> * <i>OSMR</i> *, <i>CDH13</i> *, <i>DENND1B</i> *, <i>DNMT3A</i> *, <i>FOSL2</i> *, <i>JAZF1</i> *, <i>KSR1</i> *, <i>LPP</i> *, <i>TAB2</i> *, <i>NDFIP1</i> *, <i>NFATC1</i> *, <i>PLCL1</i> *, <i>RFT1</i> *, <i>RSPO3</i> *, <i>SLAMF8</i> *, <i>THADA</i> *, <i>UBE3D</i> *, <i>ADCY3</i> * <i>BANK1</i> *, <i>BSN</i> *, <i>CDC37</i> *, <i>DAP</i> *, <i>FCGR2A</i> *, <i>HORMA2</i> *, <i>PLA2G4A</i> *, <i>SMURF1</i> *, <i>TRAF3IP2</i> *, <i>TSPAN14</i> *
Genetic evidence in treatment response	Azathioprine	<i>TPMT</i> (41), <i>ITPA</i> (4), <i>GSTA1</i> (1)
	Anti-TNF	<i>TLR2</i> (26), <i>IFNGR1</i> (2), <i>TBX21</i> (2)
	Infliximab	<i>FCGR3A</i> (11), <i>IL17F</i> + (7), <i>CASP9</i> (7), <i>ADAM17</i> + (3), <i>TRAF3IP2</i> (2), <i>FCGR1A</i> (1), <i>TNFAIP6</i> (1), <i>ZNF133</i> (1)
	Corticosteroid dependency and resistance	<i>NR3C1</i> (5)
	Thiopurine	<i>NUDT15</i> (7), <i>NAT1</i> (2)
Genetic evidence in related complications	Immunosuppressive therapy necessity	<i>IL1R1</i> (4)
	Surgical intervention	<i>HSPA2</i> (6), <i>CHRNA5</i> (1), <i>CNTF</i> (1), <i>MMP9</i> (1), <i>TSPAN14</i> (1), <i>SMURF1</i> (1)
	Structuring behavior /Aggressive disease progression	<i>SERPINE1</i> (4), <i>IDO1</i> (2), <i>SELL</i> (2), <i>HLA-DOA</i> (1), <i>CSF2RA</i> (1), <i>CYBA</i> (1), <i>FAAH</i> (1), <i>ZBTB44</i> (1), <i>CACNA1E</i> (1), <i>XPO1</i> (1), <i>KIAA1614</i> (1), <i>SULF2</i> (1)
	Food intolerance (mustard, ginger, tomatoes, wasabi)	<i>FOXO3</i> (3),
	TNF production	<i>NLRP12</i> (3)
	Granuloma formation	<i>ATG4A</i> (2), <i>ATG2A</i> (1), <i>ATG4D</i> (1), <i>FNBP1L</i> (1)
	Develop CD before 40 years of age	<i>CNR1</i> (2)
	Bone mineral density	<i>COL1A1</i> (2)
	Ileal CD	<i>FUT3</i> (2), <i>MIR196A2</i> (2), <i>MIR122</i> (1)
	Variation of GMSI level	<i>RCL1</i> (1)
	Pouch outcome	<i>DAGLB</i> (1)
	Vitamin D levels	<i>SCUBE3</i> (1), <i>PHF11</i> (1)
	Fibrostenotic CD	<i>PRPF31</i> (1)
	Favorable disease recurrence	<i>TIMP2</i> (1), <i>CYP26B1</i> (1)
	Tuberculosis and CD	<i>IL22RA1</i> (1)
	Stenotic complications	<i>MMP3</i> (1), <i>MMP1</i> (1)
	Linear growth affected	<i>DYM</i> (1)
Colon location in CD	<i>MIR124-1</i> (1)	

each module, the genes were also functionally analyzed to identify their potential phenotypic consequences. For this, we also used the Gene Network tool. Only exclusive terms for each module that were significant after Bonferroni correction were analyzed (see [Methods](#)).

For the first module (green), 60 phenotypes were retrieved; most of them related to severe symptoms such as *Ocular complications*, *Altered immune system*, *Sepsis*, *Bowel incontinence*, *Heart complications*, *Endocrine abnormalities*, *Dysphagia and constipation*, and *muscle*

problems. For the second module (purple), 11 terms were identified mainly related to *Neoplasm of the gastrointestinal tract*, *Hyperhidrosis*, *Respiratory complications*, and *Visual impairment*. These symptoms are among common abnormalities detected in IBD patients [53–55]. Finally, for the third module (blue), only four consequences were identified, *abnormality of glutamine metabolism*, *abnormality of the small intestine*, and two remaining terms related to the *facial skeleton*. The information related to the module assigned to each gene is provided in [Table S3](#).



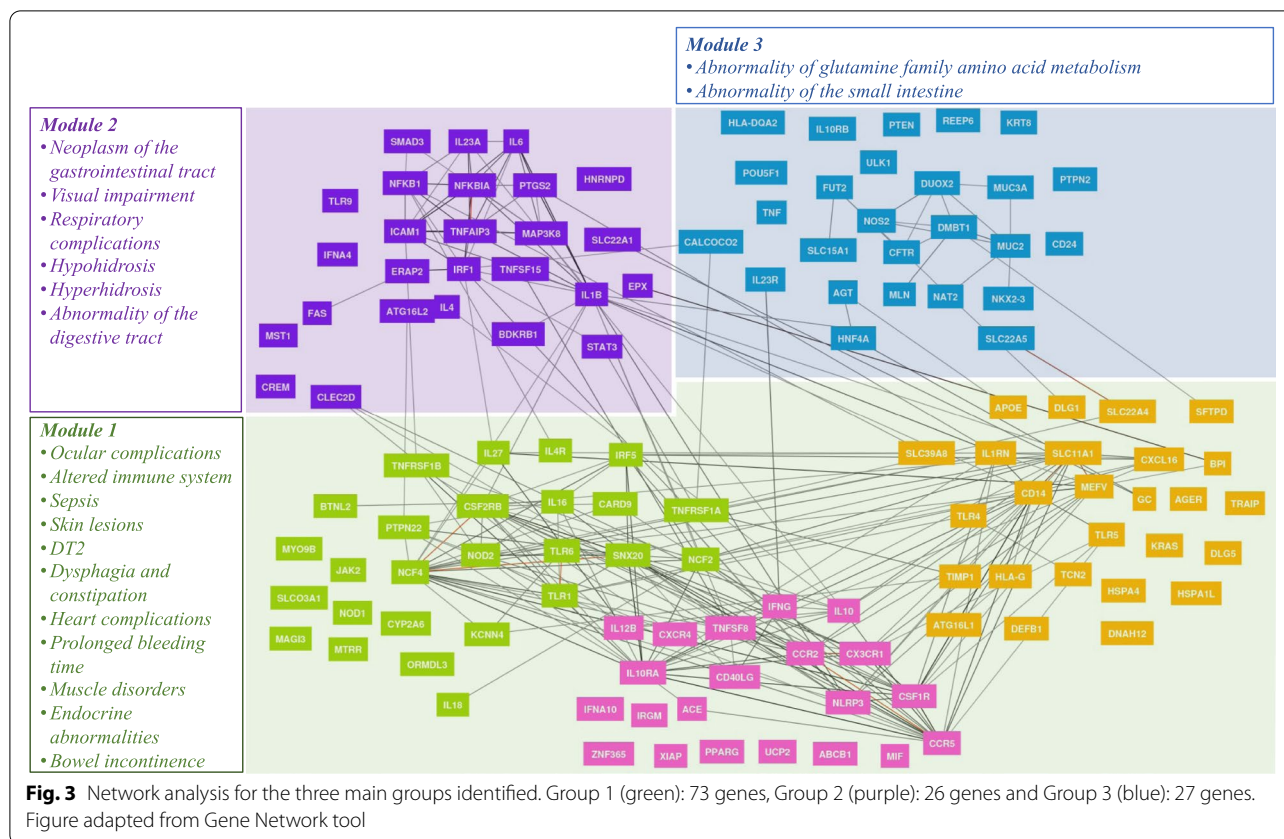
Gene expression analysis

A recent gene expression analysis of CD, UC, and controls in the colon and ileum showed that 1008 genes were differentially expressed [56]. We, therefore, explored whether the genes found genetically associated with CD in our systematic review were related to those differentially expressed (DE) between CD or UC relative to their normal ileum and colon gene expression. From the 126 genes, 67 genes were found to be DE for both UC and CD (Fig. 2B). The overlap between the 67 genes with those 1008 is highly significant ($p = 1^{-290}$, hypergeometric test), suggesting that our 126 genes are particularly enriched in DE genes. Except for *SLC39A8*, *FAS*, *IRF5*, *HSPA1L*, and *PTEN*, the vast majority were indeed more significantly associated with UC than with CD. Moreover, the majority of the genes were less expressed in CD relative to controls. We noted two solute carriers more expressed in CD than in normal colon or ileum; *SLC22A4* (ergothioneine, carnitine, tetraethylammonium) probably for detoxification, and *SLC22A5* (carnitine), whose variants are reported to affect the function of carnitine and

organic cation transporters [57]. There are also two less expressed solute carriers, *SLC11A1*, with a role in the susceptibility of humans and animals to several infections [58] and *SLC39A8*, associated with gut microbiome composition [59].

Drug-gene interaction analysis

The main therapeutic drugs for CD are azathioprine [60], infliximab, adalimumab, certolizumab pegol, ustekinumab, vedolizumab [61–63], prednisone, hydrocortisone, and hydrocortisone acetate [61]. To explore drug-gene interactions, an analysis was performed in DGIdb [41] using these drugs. There is a total of 78 genes that had a reported interaction with these CD therapeutic drugs, 10 of which were identified in our systematic review (Table 4). We reasoned that focusing on drugs that target the gene variants associated with IBD could be a strategy for CD drug repurposing [64]. Thus, to search for alternative drugs for CD, we used 13 other drugs commonly employed in the treatment of chronic autoimmune and inflammatory diseases [64, 65]. We identified 13 CD



genes (IL1B, IL1RN, IL6, ABCB1, XIAP, IFNG, ICAM1, NLRP3, JAK2, PPARG, PTGS2, APOE, and SMAD3) that show some interaction and therefore could be further explored as possible treatments for CD (Table 4).

Comparison of genetic panels

To compare the generated list of genes with genetic panels already in use, we benchmarked within those panels in the GTR [42]. We found 21 panels, of which 19 were specific to Crohn’s disease containing only 2 genes, *NOD2* and *IL6*. The two remaining tests were not specific for Crohn’s, IBD, and related diseases. These tests considered 70 genes, of which 22 were identified as functional variants for CD in our curation (including *NOD2* and *IL6*). Thus, from the 256 genes we found (Tables 3 and S3), 225 genes were not included in any panels for CD or IBD-related disorders.

We also verified the identifications at the Open Targets (OT) platform, whose pipeline includes a fine mapping of variants [31]. Filtering 3093 genes for CD for genetic association score >0.8, 178 genes were identified. Of these, 3 genes (SNN, SH2B3, and SKAP2) were not identified in the set of 1092 curated genes. From the 126 genes identified here having experimental evidence, 39 showed a low genetic association score for CD (0.5 to

0.79), 8 genes showed a good score (0.8 to 1.0) for IBD, and 52 genes do not have genetic data information in OT for CD nor IBD. Some of the 52 genes show variations that have not been reported in GWAS, explaining their absence in OT. Examples include *NOD1*, *ABCB1*, *IL1RN*, *MEFV*, and *IL18* having variants that could not be easily identified in GWAS because there are triallelic changes [66], deletions [67], and VNTR [68]. This comparison shows that even GWAS information can leave aside some information of other variants detected through other technologies or methodologies.

Discussion

Through our methodology, we have identified 256 genes associated with some aspects of CD. Of them, 126 genes were associated with *experimental evidence of variants in CD*, 71 genes found in *GWAS with a sequence variant within the gene*, 41 genes for *complications*, and 18 genes for *treatment response in CD*.

There is an explosion of genetic data provided by the high throughput technologies such as genome-wide SNP arrays and next-generation sequencing. This growing list of associated genes has the potential to improve diagnosis and treatment, but progress has been slow. There is a need for better strategies for prioritization and curation.

Table 4 Interactions between therapeutic drugs and the 126 genes

Disease	Drug	Gene interaction
Crohn's disease	Infliximab	TNF, TLR4
	Prednisone	APOE, IFNG, ABCB1
	Hydrocortisone	NOS2, ABCB1, IL1B, AGT
	Adalimumab	TNF
	Ustekinumab	IL12B, IL23A
	Certolizumab pegol	TNF
	Azathioprine	–
	Hydrocortisone acetate	–
	Vedolizumab	–
	Chronic autoimmune and inflammatory diseases	Canakinumab
Riloncept		IL1B, IL1RN
Metronidazole		IL6
Rituximab		ABCB1, XIAP
Methylprednisolone		ABCB1, IFNG
Methotrexate		IL1RN
Natalizumab		ICAM1
Anakinra		NLRP3
Olsalazine		IFNG
Tofacitinib		JAK2
Sulfasalazine		PPARG, PTGS2
Triamcinolone		APOE
Dexamethasone		SMAD3

In this study, we found that from 126 genes with variants associated with CD, a total of 110 genes have not been included in any genetic panel for CD and related diseases. That is probably reflected by the lack of individual predictive value of most individual common SNPs. The small number of variants annotated in ClinVar [32] seems to be caused by some variants not found by GWAS. Also, the increasing tendency of acquiring genetic data suggests that more efforts and more accurate annotations, such as those provided here, are highly needed and valuable.

In-silico functional analysis

The functional bioinformatics analysis performed confirmed the relationship between CD and highlighted modules of genes in our systematic review. We identified autoimmune diseases that could have affected pathways similar to those of CD such as *Type 1 Diabetes*, *Multiple sclerosis*, *Lupus*, *Arthritis Rheumatoid*, and *Psoriasis*. These relationships among CD and other autoimmune diseases are already known and have been previously studied [69, 70]. There are also relationships with hypersensitivity diseases such as asthma and celiac disease and metabolic complications such as T2D and hypercholesterolemia. Indeed, there are some Previous studies have shown an association of CD and IBD with asthma [71],

type 1 diabetes [69], and T2D [72]. Those diseases identified in our analysis are likely to share a genetic background with CD due to their inflammation process and their condition as autoimmune diseases, as suggested by previous studies [69, 71, 72]. Our results highlight the genes which could be shared among conditions and allow focusing on future research efforts among these genes.

Additionally, we found functions related to immune response, cytokine activity, and receptors. It is clear that CD pathogenesis is caused by an immune imbalance [73], which was also reflected in our de novo analysis. Some hypotheses have attempted to explain its mechanisms, including delayed hypersensitivity, activation induced by food, and others [73]. These mechanisms converge into the immune response in an environment where self-tolerance has been lost and where cytokines have an active role in maintaining this pro-inflammatory state [73, 74]. Additionally, other terms, such as apoptosis and response to Lipopolysaccharides (LPS), may provide interesting insights. LPS response is related to a monocyte/macrophage stimulation by enteric bacteria constituents [75], and resistance to apoptosis in patients with CD has also been reported [76].

We spotted signaling pathways specific for some important genes in CD converging on the activation of NF- κ B, which is a protein complex that controls the transcription of DNA cytokine production and cell survival [45]. This is comparable with previous reports of abnormal activation of NF- κ B, causing chronic inflammation in the bowel [45]. Similarly, pathways related to infections caused by protozoa, virus, and bacteria were identified consistent with the known relationship between microbes and CD [46]. Pathogen infections are one of the environmental factors which are likely to be a key component for CD; however, their roles or mechanisms of action remain speculative [77]. Additionally, the microbiota plays an important role [56]. Our results show that most of the genes related to pathogen infections are among the *common* genes and close to the pathways of *NOD*, *TLR*, and *NF κ B*, which could aid in the future understanding of the mechanisms of action specifically in CD. We also observed the pathway for *Osteoclast differentiation*, which has been recently studied, linking the function of *IL-17*, and *TNF α* modulating bone resorption [78].

In our functional analysis, we spotted the following genes, *NOD2*, *IL23R*, *IL6*, *IRGM*, *ATG16L1*, and *IL10*, whose CD-predominant risk associations are known [48, 79–82]. Among them, *NOD2* has the highest contribution to CD risk alone, with 5% of penetrance and ~20 fold risk [82]. Other genes that are not currently present in any diagnostic test for CD or a related condition but which showed general importance for CD, related diseases, and

biological process, molecular function, and pathways are *TNF*, *ICAM1*, *NFKBIA*, *NFKB1*, *TNFRSF1A*, *CD14*, *ACE*, *TLR4*, *TLR9*, *IFNG*, *IL1B*, *IL4*, *IL12B*, *IL1RN*, *IL18*, and *IL4R*, which are involved in TNF and NF- κ B signaling pathways, in inflammatory response, and cytokine interaction processes [47–52]. These genes and others from our list could be used to design a more robust prediction panel for CD risk.

Our analysis highlighted poorly studied genes (10 or fewer abstracts). From these, *FUT2*, *DNAH12*, *TRAIIP*, and *ERAP1* were identified in the functional analysis to be specific for CD (Fig. 2). Among the processes reported for these genes are ABH antigens expression [83], motile cilia function [84], regulation of innate immune signaling [85], and immune activation and inflammation [86].

Among these poorly studied genes, only *FUT2* is currently present in a diagnostic panel related to IBD diseases. This fact remarks the importance of considering and further studying the biological implication of the less studied set of genes to increase our knowledge of this complex disease.

Network analysis

We identified three network modules of genes associated with specific symptoms. The first module comprising 73 genes was related to severe symptoms [87], such as *Altered immune system*, *Sepsis*, *Bleeding*, *Muscle disorders*, and *Heart complications* well-known in CD. The second module involved 26 genes related to *hyperhidrosis*, *respiratory complications*, and *neoplasm of the gastrointestinal tract*. These symptoms are among the common abnormalities detected in IBD patients [53–55]. The third module, including 26 genes, was associated with *abnormality of glutamine metabolism* and *abnormality of the small intestine*. Glutamine is an important supplementation in IBD patients [88], and its effects in IBD have been studied in animal models [89] and patients [90, 91]. Thus, this module seems to map genes related to less severe consequences for CD. Thus, gene-symptom mapping may provide important insights into CD.

Gene expression analysis

Differential expression analysis of the 126 genes identified in our systematic review revealed that a significant number of genes show dysregulated levels of expression in colon and ileum biopsies of both CD and UC when compared with not-IBD patients further supporting our gene prioritization approach. The greatest changes in expression are observed in the colon, with differential expression in pro-inflammatory genes (*NOD2*, *IL1B*, and *TNF*). Other observations are that changes are different among UC and CD and that a large proportion of the genes do not show evident gene

expression. Thus, to further understand whether the functional implications of these changes in expression are causal for CD pathogenesis or whether CD patients carrying other specific variants show different gene expression profiles, further functional experiments are needed.

Drug-gene interactions

Among the 126 genes analyzed, 10 have a reported interaction with known CD therapeutic drugs, and 13 have a reported interaction with other autoimmune and inflammatory diseases. The treatment for CD is complex, and it is focused on controlling the symptoms and the remission of the disease [92]. Focusing on drugs that target the gene variants associated with IBD can be a strategy for CD drug development [64]. This highlights the necessity of considering more genes to study other possible interactions for CD beyond what is currently known and shows the importance of gene curation strategies, like the one proposed here. Further research on the genes highlighted here, and their mechanisms of interaction with CD diseases could improve the knowledge of the disease development and expand treatments.

Traditional drug development is costly and can take 10–15 years to develop an efficient drug [64]. Personalized medicine exhibits the clinical application of drug-gene interaction, where drugs are guided based on the individual's genetics and disease progress. Targeting CD's genetic risk regions that had been experimentally validated can improve the identification of possible drug candidates. This can be reflected in target-directed therapies, which is one of the main objectives of personalized medicine. The analysis of drug-gene interactions in a complex disease, such as MDD (major depressive disorder), allowed a better, prioritization of drug-genes sets and the identification of drugs indicating an effect on a disease, reflecting potential repurposing opportunities [93]. Nevertheless, validation studies are still required to ensure the drug-gene interaction and avoid side effects.

Our results support the consideration of several genes when studying CD. More importantly, the functional analysis provides a mapping between genes and key aspects of Crohn's disease. The integration of other genes may also be important. For example, genes close by a non-coding GWAS SNP, i.e., intergenic variants, or those involved in related diseases, could play a role in CD etiology, but further validation or fine gene mapping is needed.

Abbreviations

CD: Crohn's disease; IBD: Inflammatory Bowel Disease; UC: Ulcerative Colitis; GWAS: Genome-Wide Association Studies; SNP: Single Nucleotide Polymorphism; NGS: Next-Generation Sequencing; PRS: Polygenic Risk Score; GAD: Genetic Association Database; GO: Gene Ontology; HPO: Human Phenotype Ontology; DGIdb: Drug-Gene Interaction Database; GTR: Genetic Test Registry.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08491-y>.

Additional file 1. PRISMA checklist.

Additional file 2. List of Pubmed IDs for genes curated through PubTerm.

Additional file 3. List of Pubmed IDs for genes curated from GWAS Catalog.

Additional file 4: Table S1. Annotations for more than 1000 genes for their association to CD.

Additional file 5: Table S2. Annotations for 133 genes from GWAS Catalog associated to CD.

Additional file 6: Table S3. Details of the 126 genes categorized as experimental evidence for CD.

Additional file 7: Table S4. Mutations of associated to genes not annotated in other databases.

Additional file 8: Table S5. Annotation for genes categorized as other genetic associations.

Acknowledgments

Not applicable.

Authors' contributions

DGH and VT designed the study. DGH performed the functional analysis and wrote the initial version of the manuscript. MSV and VT were major contributors in writing the manuscript. VT supervised the study. JGP aided in gene curation. RAG aided in the differential expression analysis. PL, MMV, and KE reviewed the results and provided suggestions and recommendations. All authors read and approved the final manuscript.

Funding

DGH received a scholarship for PhD studies from CONACyT (student id 501329).

Availability of data and materials

The prioritized gene catalog can be explored at <http://victortrevino.bioinformatics.mx/CrohnDisease>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Tecnologico de Monterrey, Escuela de Medicina, Cátedra de Bioinformática, Av. Morones Prieto No. 3000, Colonia Los Doctores, 64710 Monterrey, Nuevo León, Mexico. ²Instituto de Investigação e Inovação em Saúde-i3S, Universidade do Porto, Porto, Portugal. ³Ipatimup, Institute of Molecular Pathology and Immunology at the University of Porto, Porto, Portugal. ⁴Department of Human Genetics, University of Chicago, Chicago, USA. ⁵McGill University Health Centre, Division of Gastroenterology, IBD Centre, Montreal General

Hospital, 1650 Ave. Cedar, D16.173.1, Montreal, QC H3G 1A4, Canada. ⁶Graduate Professional Studies, Brandeis University, Waltham, MA 02453, USA. ⁷Tecnologico de Monterrey, Instituto de Medicina Interna, Centro Médico Zambrano Hellion, Av. Batallón de San Patricio No. 112, Colonia Real San Agustín, 66278 San Pedro Garza García, Nuevo León, Mexico. ⁸Tecnologico de Monterrey, The Institute for Obesity Research, Integrative Biology Unit, Eugenio Garza Sada 2501 Avenue, 64849 Monterrey, Nuevo Leon, Mexico.

Received: 5 October 2021 Accepted: 28 February 2022

Published online: 13 April 2022

References

- Pia Costa Santos M, Gomes C, Torres hospital Beatriz Ângelo J. familial and ethnic risk in inflammatory bowel disease. *Ann Gastroenterol.* 2018;31:14–23.
- Baumgart DC, Sandborn WJ. Crohn's disease. *Lancet.* 2012;380:1590–605.
- Yaari S, Benson A, Aviran E, Lev Cohain N, Oren R, Sosna J, et al. Factors associated with surgery in patients with intra-abdominal fistulizing Crohn's disease. *World J Gastroenterol.* 2016;22:10380–7.
- Feuerstein JD, Cheifetz AS. Crohn disease: epidemiology, diagnosis, and management. *Mayo Clin Proc.* 2017;92:1088–103.
- Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet.* 2017;390:2769–78.
- Ahmad T, Satsangi J, MCGovern D, Bunce M, Jewell DP. The genetics of inflammatory bowel disease. *Aliment Pharmacol Ther.* 2001;15:731–48.
- Gordon H, Moller FT, Andersen V, Harbord M. Heritability in inflammatory bowel disease: from the first twin study to genome-wide association studies. *Inflamm Bowel Dis.* 2015;21:1428–34.
- Ellinghaus D, Bethune J, Petersen B-S, Franke A. The genetics of Crohn's disease and ulcerative colitis – *status quo* and beyond. *Scand J Gastroenterol.* 2015;50:13–23.
- Cleynen I, Boucher G, Jostins L, Schumm LP, Zeissig S, Ahmad T, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet (London, England).* 2016;387:156–67.
- Liu JZ, Anderson CA. Genetic studies of Crohn's disease: past, present and future. *Best Pract Res Clin Gastroenterol.* 2014;28:373–86.
- Gajendran M, Loganathan P, Catinella AP, Hashash JG. A comprehensive review and update on Crohn's disease. *Disease-a-Month.* 2018;64:20–57.
- Michail S, Bultron G, Depaolo RW. Genetic variants associated with Crohn's disease. *Appl Clin Genet.* 2013;6:25–32.
- Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. *Nat Rev Genet.* 2013;14:415–26.
- Reid ES, Papatreou A, Drury S, Boustred C, Yue WW, Wedatilake Y, et al. Advantages and pitfalls of an extended gene panel for investigating complex neurometabolic phenotypes. *Brain.* 2016;139:2844–54.
- Biasci D, Lee JC, Noor NM, Pombal DR, Hou M, Lewis N, et al. A blood-based prognostic biomarker in IBD. *Gut.* 2019;68:1386–95.
- Mascheretti S, Schreiber S. Genetic testing in Crohn disease: utility in individualizing patient management. *Am J Pharmacogenomics.* 2005;5:213–22.
- de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet.* 2017;49:256–61.
- Liu JZ, Van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet.* 2015;47:979–86.
- Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet.* 2008;40:955–62.
- Chen G-B, Lee SH, Montgomery GW, Wray NR, Visscher PM, Geary RB, et al. Performance of risk prediction for inflammatory bowel disease based on genotyping platform and genomic risk score method. *BMC Med Genet.* 2017;18:94.

21. Moninuola OO, Milligan W, Lochhead P, Khalili H. Systematic review with meta-analysis: association between acetaminophen and nonsteroidal anti-inflammatory drugs (NSAIDs) and risk of Crohn's disease and ulcerative colitis exacerbation. *Aliment Pharmacol Ther.* 2018;47:1428–39.
22. Barnes EL, Kochar B, Jessup HR, Herfarth HH. The incidence and definition of Crohn's disease of the pouch: a systematic review and Meta-analysis. *Inflamm Bowel Dis.* 25(9):1474–80.
23. Karoui S, Serghini M, Dachraoui A, Boubaker J, Filali A. Prognostic factors in Crohn's disease: a systematic review. *Tunis Med.* 2013;91:230–3.
24. Cui G, Yuan A. A systematic review of epidemiology and risk factors associated with Chinese inflammatory bowel disease. *Front Med.* 2018;5:183.
25. Peng Q-H, Wang Y-F, He M-Q, Zhang C, Tang Q. Clinical literature review of 1858 Crohn's disease cases requiring surgery in China. *World J Gastroenterol.* 2015;21:4735–43.
26. Garcia-Rivas G, Jerjes-Sánchez C, Rodriguez D, Garcia-Pelaez J, Trevino V. A systematic review of genetic mutations in pulmonary arterial hypertension. *BMC Med Genet.* 2017;18:82.
27. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *PLoS Med.* 2021;18:e1003583.
28. Garcia-Pelaez J, Rodriguez D, Medina-Molina R, Garcia-Rivas G, Jerjes-Sánchez C, Trevino V. PubTerm: a web tool for organizing, annotating and curating genes, diseases, molecules and other concepts from PubMed records. *Database (Oxford).* 2019;2019:bay137.
29. Sepulveda-Villegas M, Elizondo-Montemayor L, Trevino V. Identification and analysis of 35 genes associated with vitamin D deficiency: a systematic review to identify genetic variants. *J Steroid Biochem Mol Biol.* 2020;196:105516.
30. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47:D1005–12.
31. Carvalho-Silva D, Pierleoni A, Pignatelli M, Ong C, Fumis L, Karamanis N, et al. Open targets platform: new developments and updates two years on. *Nucleic Acids Res.* 2019;47(D1):D1056–D1065.
32. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46:D1062–7.
33. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4:44–57.
34. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat Genet.* 2000;25:25–9.
35. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 2019;47:D330–8.
36. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database [1]. *Nat Genet.* 2004;36:431–2.
37. Deelen P, van Dam S, Herkert JC, Karjalainen JM, Abbott KM, van Diemen CC, et al. Improving the diagnostic yield of exome-sequencing by predicting gene-phenotype associations using large-scale gene expression analysis. *Nat Commun.* 2019;10:2837.
38. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gouridine J-P, et al. Expansion of the human phenotype ontology (HPO) knowledge base and resources. *Nucleic Acids Res.* 2019;47:D1018–27.
39. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets - update. *Nucleic Acids Res.* 2013;41:D991–5.
40. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* 1995.pdf. *J R Stat Soc Ser B.* 1995;57:289–300.
41. Cotto KC, Wagner AH, Feng Y-Y, Kiwala S, Coffman AC, Spies G, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* 2018;46:D1068–73.
42. Rubinstein WS, Maglott DR, Lee JM, Kattman BL, Malheiro AJ, Ovetsky M, et al. The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive information and improve transparency. *Nucleic Acids Res.* 2013;41(D1):D925–D935.
43. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29:308–11.
44. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37:1–13.
45. Ellis RD, Goodlad JR, Limb GA, Powell JJ, Thompson RPH, Pouchard NA. Activation of nuclear factor kappa B in Crohn's disease. *Inflamm Res.* 1998;47:440–5.
46. De Hertogh G, Aerssens J, Geboes KP, Geboes K. Evidence for the involvement of infectious agents in the pathogenesis of Crohn's disease. *World J Gastroenterol.* 2008;14:845–52.
47. García-Ramírez RA, Ramírez-Venegas A, Quintana-Carrillo R, Camarena ÁE, Falfán-Valencia R, Mejía-Aranguré JM. TNF, IL6, and IL1B polymorphisms are associated with severe influenza A (H1N1) virus infection in the Mexican population. *PLoS One.* 2015;10:e0144832.
48. Turner MD, Nedjai B, Hurst T, Pennington DJ. Cytokines and chemokines: at the crossroads of cell signalling and inflammatory disease. *Biochim Biophys Acta - Mol Cell Res.* 2014;1843:2563–82.
49. Zhang G-L, Zou Y-F, Feng X-L, Shi H-J, Du X-F, Shao M-H, et al. Association of the NFKBIA gene polymorphisms with susceptibility to autoimmune and inflammatory diseases: a meta-analysis. *Inflamm Res.* 2011;60:11–8.
50. Kopitar-Jerala N. The role of interferons in inflammation and Inflammation activation. *Front Immunol.* 2017;8:873.
51. Gurung P, Malireddi RKS, Anand PK, Demon D, Vande Walle L, Liu Z, et al. Toll or interleukin-1 receptor (TIR) domain-containing adaptor inducing interferon-β (TRIF)-mediated caspase-11 protease production integrates toll-like receptor 4 (TLR4) protein- and Nlrp3 inflammasome-mediated host defense against enteropathogens. *J Biol Chem.* 2012;287:34474–83.
52. Glas J, Seiderer J, Wagner J, Olszak T, Fries C, Tillack C, et al. Analysis of IL12B gene variants in inflammatory bowel disease. *PLoS One.* 2012;7:e34349.
53. Davidson JRT, Foa EB, Connor KM, Churchill LE. Hyperhidrosis in social anxiety disorder. *Prog Neuro-Psychopharmacol Biol Psychiatry.* 2002;26:1327–31.
54. Bannaga AS, Selinger CP. Inflammatory bowel disease and anxiety: links, risks, and challenges faced. *Clin Exp Gastroenterol.* 2015;8:111–7.
55. Majewski S, Piotrowski W. Pulmonary manifestations of inflammatory bowel disease. *Arch Med Sci.* 2015;11:1179–88.
56. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature.* 2019;569:655–62.
57. Sartor RB. Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nat Clin Pract Gastroenterol Hepatol.* 2006;3:390–407.
58. Sechi LA, Dow CT. *Mycobacterium avium* ss. paratuberculosis Zoonosis - The Hundred Year War - Beyond Crohn's Disease. *Front Immunol.* 2015;6:96.
59. Li D, Achkar JP, Haritunians T, Jacobs JP, Hui KY, D'Amato M, et al. A pleiotropic missense variant in SLC39A8 is associated with Crohn's disease and human gut microbiome composition. *Gastroenterology.* 2016;151:724–32.
60. Magro F, Santos-Antunes J, Vilas-Boas F, Rodrigues-Pinto E, Coelho R, Ribeiro OS, et al. Crohn's disease outcome in patients under azathioprine: a tertiary referral center experience. *J Crohn's Colitis.* 2014;8:617–25.
61. Rappaport N, Nativ N, Stelzer G, Twik M, Guan-Golan Y, Stein TI, et al. MalaCards: an integrated compendium for diseases and their annotation. *Database (Oxford).* 2013;2013:bat018.
62. Močko P, Kawalec P, Pilc A. Safety profile of biologic drugs in the therapy of Crohn disease: a systematic review and network meta-analysis. *Pharmacol Reports.* 2016;68:1237–43.
63. Flamant M, Roblin X. Inflammatory bowel disease: towards a personalized medicine. *Ther Adv Gastroenterol.* 2018;11:1–15.
64. Grenier L, Hu P. Computational drug repurposing for inflammatory bowel disease using genetic information. *Comput Struct Biotechnol J.* 2019;17:127–35.
65. Dinarello CA. Anti-inflammatory Agents: Present and Future. *Cell.* 2010;140(6):935–50.
66. Potočník U, Ferkoj I, Glavač D, Dean M. Polymorphisms in multidrug resistance 1 (MDR1) gene are associated with refractory Crohn disease and ulcerative colitis. *Genes Immun.* 2004;5:530–9.

67. MCGovern DPB, Hysi P, Ahmad T, Van Heel DA, Moffatt MF, Carey A, et al. Association between a complex insertion/deletion polymorphism in NOD1 (CARD4) and susceptibility to inflammatory bowel disease. *Hum Mol Genet.* 2005;14(10):1245–50.
68. Stankovic B, Dragasevic S, Popovic D, Zukic B, Kotur N, Sokic-Milutinovic A, et al. Variations in inflammatory genes as molecular markers for prediction of inflammatory bowel disease occurrence. *J Dig Dis.* 2015;16:723–33.
69. Benchimol EI, Manuel DG, To T, Mack DR, Nguyen GC, Gommerman JL, et al. Asthma, type 1 and type 2 diabetes mellitus, and inflammatory bowel disease amongst south Asian immigrants to Canada and their children: a population-based cohort study. *PLoS One.* 2015;10:e0123599.
70. Kosmidou M, Katsanos AH, Katsanos KH, Kyritsis AP, Tsivgoulis G, Christodoulou D, et al. Multiple sclerosis and inflammatory bowel diseases: a systematic review and meta-analysis. *J Neurol.* 2017;264:254–9.
71. Kuenzig ME, Bishay K, Leigh R, Kaplan GG, Benchimol EI, Crowdscreen SR, et al. Co-occurrence of asthma and the inflammatory bowel diseases: A Systematic Review and Meta-analysis. *Clin Transl Gastroenterol.* 2018;9:188.
72. Jurjus A, Eid A, Al Kattar S, Zeenny MN, Gerges-Geagea A, Haydar H, et al. Inflammatory bowel disease, colorectal cancer and type 2 diabetes mellitus: the links. *BBA Clin.* 2016;5:16–24.
73. Li N, Shi R-H. Updated review on immune factors in pathogenesis of Crohn's disease. *World J Gastroenterol.* 2018;24:15–22.
74. Segal AW. Making sense of the cause of Crohn's - a new look at an old disease. *F1000Research.* 2016;5:2510.
75. Caradonna L, Amati L, Magrone T, Pellegrino NM, Jirillo E, Caccavo D. Enteric bacteria, lipopolysaccharides and related cytokines in inflammatory bowel disease: biological and clinical significance. *J Endotoxin Res.* 2000;6:205–14.
76. Eder P, Łykowska-Szuber L, Krela-Kaźmierczak I, Stawczyk-Eder K, Iwanik K, Majewski P, et al. Disturbances in apoptosis of lamina propria lymphocytes in Crohn's disease. *Arch Med Sci.* 2015;11:1279–85.
77. Hubbard VM, Cadwell K. Viruses, autophagy genes, and Crohn's disease. *Viruses.* 2011;3:1281–311.
78. Ciucci T, Ibáñez L, Boucoiran A, Birgy-Barelli E, Pène J, Abou-Ezzi G, et al. Bone marrow Th17 TNF α cells induce osteoclast differentiation, and link bone destruction to IBD. *Gut.* 2015;64:1072–81.
79. Chauhan S, Mandell MA, Deretic V. IRGM governs the core autophagy machinery to conduct antimicrobial defense. *Mol Cell.* 2015;58:507–21.
80. Travassos LH, Carneiro LAM, Ramjeet M, Hussey S, Kim Y-G, Magalhães JG, et al. Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nat Immunol.* 2010;11:55–62.
81. McCarroll SA, Huett A, Kuballa P, Chilewski SD, Landry A, Goyette P, et al. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet.* 2008;40:1107–12.
82. Cho JH, Brant SR. Recent insights into the genetics of inflammatory bowel disease. *Gastroenterology.* 2011;140:1704–12.
83. Wacklin P, Mäkiyuokko H, Alakulppi N, Nikkilä J, Tenkanen H, Räsänen J, et al. Secretor genotype (FUT2 gene) is strongly associated with the composition of bifidobacteria in the human intestine. *PLoS One.* 2011;6:e20113.
84. Miller M, Tam AB, Cho JY, Doherty TA, Pham A, Khorram N, et al. ORMDL3 is an inducible lung epithelial gene regulating metalloproteases, chemokines, OAS, and ATF6. *Proc Natl Acad Sci U S A.* 2012;109:16648–53.
85. Hoffmann S, Smedegaard S, Nakamura K, Mortuza GB, Räschle M, Ibáñez de Opakua A, et al. TRAP1 is a PCNA-binding ubiquitin ligase that protects genome stability after replication stress. *J Cell Biol.* 2016;212:63–75.
86. Paladini F, Fiorillo MT, Vitulano C, Tedeschi V, Piga M, Cauli A, et al. An allelic variant in the intergenic region between ERAP1 and ERAP2 correlates with an inverse expression of the two genes. *Sci Rep.* 2018;8:10398.
87. Mantzaris GJ, Viazis N, Polymeros D, Papamichael K, Bamias G, Koutroubakis IE. Clinical profiles of moderate and severe crohn's disease patients and use of anti-tumor necrosis factor agents: Greek expert consensus guidelines. *Ann Gastroenterol.* 2015;28:417–25.
88. Kim M-H, Kim H. The roles of glutamine in the intestine and its implication in intestinal diseases. *Int J Mol Sci.* 2017;18(5):1051.
89. Fillmann H, Kretzmann NA, San-Miguel B, Llesuy S, Marroni N, González-Gallego J, et al. Glutamine inhibits over-expression of pro-inflammatory genes and down-regulates the nuclear factor kappaB pathway in an experimental model of colitis in the rat. *Toxicology.* 2007;236:217–26.
90. Akobeng AK, Elawad M, Gordon M. Glutamine for induction of remission in Crohn's disease. *Cochrane Database Syst Rev.* 2016;2:CD007348.
91. Liu Y, Wang X, Hu C-AA. Therapeutic potential of amino acids in inflammatory bowel disease. *Nutrients.* 2017;9:920.
92. Rufini S, Ciccacci C, Novelli G, Borgiani P. Pharmacogenetics of inflammatory bowel disease: a focus on Crohn's disease. *Pharmacogenomics.* 2017;18:1095–114.
93. Gaspar HA, Gerring Z, Hübel C, Middeldorp CM, Derks EM, Breen G. Using genetic drug-target networks to develop new drug hypotheses for major depressive disorder. *Transl Psychiatry.* 2019;9:1–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

