

**A GUIDE TO USING THE
INFLAMMATORY BOWEL DISEASE
DATABASE
(IBDDB)**

OCTOBER 2020

Table of contents

A. Introduction.....	3
B. Accessing IBDDDB	4
C. Exploration of list of reports generated.....	5
1. IBD database.....	5
2. Knowledgebase	6
3. Biomedical entities.....	7
4. Biomedical entities co-occurrence	8
5. Hypotheses explorer	9
D. Network creation and visualization.....	10
E. Downloading Data.....	11

A. Introduction

The Inflammatory Bowel Disease Database (IBDDDB) is the first manually curated database of experimentally verified IBD genes. This database provides pre-compiled biomedical text-mining of PubMed (manually cleaned and curated) and PubChem abstracts. The various information embedded in the database is viewed in IBDDDB via entities identified as a subject-specific set of documents. These entities are terms, assembled into classified dictionaries.

IBDDDB exhibits 11 specialised dictionaries:

- Biological Action
- Biological Processes
- Cellular Component
- ChEBI
- Dietary Supplement
- Diseases
- Drugs
- Human Genes
- Metabolites-and-Enzymes
- Pathways
- Toxins

The user can also explore different biomedical entities (about 20,000) and their co-occurrence with other entities (about one million) from 11 dictionaries, which are color-coded, in the indexed PubMed records.

IBDDDB database uses various specialized dictionaries that are curated manually. It is important for users to contemplate that these entities are terms, assembled into classified dictionaries. For example, the 'biological process' is a dictionary, and its medical entity is 'amino acid biosynthesis', 'amino acid transport', etcetera. Meanwhile during the use of 'Disease' dictionary, one should know that this dictionary use has also emphasized on broad spectrum. This dictionary not only contains disease names, but also contain symptoms and various other related terms.

B. Accessing IBDDDB

IBDDDB allows user to explore, dissect and retrieve information using “Explore” tab with ease through an interactive user-friendly interface. By clicking on this tab, it will take to drop down menu displaying several lists of reports discussed below.

The Explore tabs containing list of reports:

- IBD database
- Knowledgebase
- Biomedical entities
- Biomedical entities co-occurrence
- Hypotheses explorer

C. Exploration of list of reports generated

1. IBD database

IBD database is a unique resource as it is a combination of curated and text-mined information which is easily explorable and user-friendly. This is the first manually curated database providing extended information about experimentally validated IBD related genes. It integrates data on 34 subject-specific concepts (listed as columns) as shown in Fig 2.

IBD Home Explore Visualize About

Inflammatory Bowel Disease (IBD) Database

show/hide columns

id	HGNC gene symbol	HGNC gene name	chromosome_location	snp_variants	IBD phenotypes	previous_name	omim_id
omim_diseases	uniprotkb	entrez_id	HGNC_id	ensembl_id	refseq_DNA_sequence	pmid	pmcid
experimental_evidences	study_subject	up_down_regulation	inflamed_sites	tissues_samples	cell_lines	literature_disease	biological_process
cellular_process	molecular_function	kegg_pathways	reactome_pathways	DGldb_interactiontypes(sources)	dsigdb	tfs_transfac	tfs_chea
tfs_encode				tfs_ompossum			

Q

Copy Excel CSV PDF Print

HGNC gene symbol	HGNC gene name	IBD phenotypes	pmid	experimental evidences	study subject
NKX2-3	NK2 homeobox 3	UC	20957151, 21072187, 20165982	qPCR, RT-PCR, Genotyping, Western blot analysis, Ingenuity Pathway analysis, Microarray and data analysis, Immunofluorescence assay	
PSMD8	Proteasome 26S subunit, Non-ATPase 8	CD	24223725	siRNA transfection, qPCR, Western blot, Electrophoretic mobility shift and Luciferase assays	Mice
NLRP3	NLR family pyrin domain containing 3	UC, CD	19784369, 20385749, 30583612, 29102545, 28523573, 27966619	SNP genotyping, qPCR, ELISA	Mice and Human

Figure 2: Explore the IBD database.

The information of IBD database page under the Explore tab integrates the data on 34 concepts such as experimental techniques used to validate the role of a gene in IBD, sites inflamed in IBD, other disease linked to the genes, tissue samples used for validation, molecular interactions, gene expression level in mice or humans, and transcription factors of IBD implicated genes using four different tools, and also contains data on drugs associated with IBD genes. The result is downloadable under various formats highlighted with red box.

2. Knowledgebase

The in-house text-mining Dragon Exploration System (DES), which has been used previously for the creation of several published knowledgebases was used to obtain a list of potential IBD-related genes. The PubMed was searched for relevant literature using queries ‘Inflammatory AND bowel AND IBD’, ‘Crohn's AND CD’, and ‘Ulcerative AND colitis AND UC’. The titles and abstracts were downloaded and indexed using several dictionaries as shown in Fig 3. We have text-mined the abstract for more than 700,000 articles and partly done manually along with their PubMed identifiers.

The screenshot displays the IBD Knowledgebase interface. At the top, there is a navigation menu with 'IBD', 'Home', 'Explore', 'Visualize', and 'About'. Below the menu is a search bar with a magnifying glass icon. A toolbar contains options for 'Copy', 'Excel', 'CSV', 'PDF', and 'Print'. A 'Dictionary selection' dropdown menu is visible. The main content area shows a list of indexed documents. Each row includes a 'Statement' column and a 'PubMed' column with a document icon. The text in the 'Statement' column contains highlighted terms in purple boxes, such as 'IBD', 'inflammatory bowel disease', and 'gastrointestinal'. At the bottom, there is a pagination control showing 'Showing 1 to 10 of 278,501 entries' and a set of buttons for 'Previous', '1', '2', '3', '4', '5', '...', '27851', and 'Next'.

Statement	PubMed
11 10.2% were diagnosed with IBD, 7.3% with another GI condition associated with an abnormal GI tract and 63.2% had functional GI disease.	
13 105 outpatients with IBD of a University Hospital agreed to participate over a 4 week follow-up period.	
14 107 of these hospitals were classified as high-volume centers (HVCs) for pediatric IBD care and 553 low-volume centers (LVCs).	
15 10% of patients with ulcerative colitis were colectomized within 1 year and 23% within 10 years.	
18 111Indium (111In) WBC scintigraphy is an accurate method of assessing the extent of inflammatory bowel disease.	
19 111Indium autologous granulocytes in the detection of inflammatory bowel disease.	
20 111Indium autologous leucocytes in inflammatory bowel disease.	
21 111In leucocyte scintigraphy always demonstrated inflamed areas of bowel, but underestimated the extent of disease in three patients and overestimated it in five.	
22 111In-oxine-labelled autologous leucocytes in inflammatory bowel disease- new scintigraphic activity index.	
23 112 journals published articles on gastrointestinal microbiome.	

Figure 3: This figure illustrates an example of indexed document for the highlighted term (on the left-side) mentioned in the PubMed literature (on the right-side).

3. Biomedical entities

It is important for users to contemplate that these entities are terms, assembled into classified dictionaries. One medical entity can appear in the various dictionary; therefore, it can be normalized by using different names from the dictionary's list (e.g. cellular component, drug agent, etc.). The user can also explore different biomedical entities (about 20,000) from 11 dictionaries in the indexed PubMed records as shown in Fig 4..

Term	Category	Freq.	PubMed
Inflammatory bowel disease (IBD) Inflammatory Bowel Disease (IBD) inflammatory bowel disease (IBD) Inflammatory Bowel Disease inflammatory bowel disease inflammatory bowel disease IBD (IBD) Inflammatory bowel disease (I.B.D.) inflammatory bowel disease (I.B.D.) (I.B.D.)	Diseases	42478	PubMed
Crohn's disease (Crohn's disease) Crohn's disease (CD) Crohn Disease Crohns Disease CD	Diseases	33003	PubMed
Ulcerative colitis (UC) ulcerative colitis (UC) Ulcerative colitis ulcerative colitis UC (UC)	Diseases	31760	PubMed
Primethamine Chloridyn Chloridine CD 5-(4-chlorophenyl)-6-ethylpyrimidine-2,4-diamine pyrimethaminum pirimetamina Ethylpyrimidine Diaminopyritamin 5-(4-Chlorophenyl)-6-ethyl-2,4-pyrimidinediamine 5-(4-Chlorophenyl)-6-ethyl-2,4-diaminopyrimidine 5-(4'-Chlorophenyl)-2,4-diamino-6-ethylpyrimidine 2,4-Diamino-5-chlorophenyl-6-ethylpyrimidine 2,4-Diamino-5-(p-chlorophenyl)-6-ethylpyrimidine 2,4-Diamino-5-(4-chlorophenyl)-6-ethylpyrimidine pyrimethamine pyrimethamine	ChEBI	10914	PubMed
corticoides corticosteroides corticosteroids	ChEBI	2294	PubMed
sodium	Metabolites-and-Enzymes	2267	PubMed
immune response	Biologica Processes	2144	PubMed
Steroid a steroid steroids	ChEBI	1997	PubMed
(4)He(2+) alpha alpha-particle alpha-particle helium-4(2+)	ChEBI	1978	PubMed

Figure 4: This figure illustrates the biomedical entities under category along with frequency of that occurring in PubMed literatures.

4. Biomedical entities co-occurrence

IBDDDB helps user to Explore co-occurrence of biomedical entities linked to IBD. It allows to find co-mention of two entities in the published literature by selecting any two dictionaries (out of 11 listed). The relevant PubMed articles can also be retrieved for further exploration as shown in Fig 5.

IBD Home **Explore** Visualize About

Explore co-occurrence of biomedical entities linked to inflammatory bowel disease (IBD)

Copy Excel CSV PDF Print

All All

Term A	Category A	Term B	Category B	Freq.	PubMed
Leukemia, Lymphocytic, Chronic, B-Cell	Diseases	Crohn's disease	Diseases	3	PubMed
Leukemia, Lymphocytic, Chronic, B-Cell	Diseases	Inflammatory bowel disease (IBD)	Diseases	2	PubMed
Leukemia, Lymphocytic, Chronic, B-Cell	Diseases	Ulcerative colitis (UC)	Diseases	2	PubMed
Leukemia, Lymphocytic, Chronic, B-Cell	Diseases	Vedolizumab	Drugs	2	PubMed
Leukemia, Lymphocytic, Chronic, B-Cell	Diseases	ALB	Human Genes	2	PubMed
Leukemia, Lymphocytic, Chronic, B-Cell	Diseases	CRP	Human Genes	1	PubMed
Leukemia, Lymphocytic, Chronic, B-Cell	Diseases	Bupropion	Metabolites-and-Enzymes	1	PubMed
Leukemia, Lymphocytic, Chronic, B-Cell	Diseases	cell adhesion	Biologica Processes	1	PubMed
Leukemia, Lymphocytic, Chronic, B-Cell	Diseases	Apoptosis	Biologica Processes	1	PubMed
Lymphoma, Non-Hodgkin	Diseases	Crohn's disease	Diseases	1	PubMed
Lymphoma, Non-Hodgkin	Diseases	Inflammatory bowel disease (IBD)	Diseases	2	PubMed
Lymphoma, Non-Hodgkin	Diseases	Ulcerative colitis (UC)	Diseases	3	PubMed
Lymphoma, Non-Hodgkin	Diseases	NDUFB6	Human Genes	1	PubMed

Page size: 25

Showing 1 to 25 of 295,023 entries

Previous 1 2 3 4 5 ... 11801 Next

Figure 5: Explore Co-occurrence of biomedical entities linked to IBD. Here terms from Term A and B are co-occurring using biomedical entities from category A and B as denoted by the frequency column. PubMed abstract can be retrieved for these co-occurring terms.

5. Hypotheses explorer

Hypotheses generator identifies the interaction between different identities from various dictionaries and their co-occurrence frequencies. Hypotheses explorer (under Explore tab) allows to identify new association among selected terms selected from different dictionaries (A-B and B-C are known and A-C are suggested links between biomedical entities) as shown in Fig 6. Users can accept or reject the abstracts based on their active association relationship manually.

The screenshot shows the 'Hypotheses explorer' interface. At the top, there is a navigation bar with 'IBD', 'Home', 'Explore' (highlighted), 'Visualize', and 'About'. Below the navigation bar, the title reads 'Hypotheses explorer (A-B and B-C are known and A-C are suggested links between biomedical entities)'. The interface is divided into three columns, each representing a different dictionary or category. The first column, 'All dictionaries', has a search box containing 'pyoderma' and a table with 'id 1: Term A' (12567) and 'Pyoderma Gangrenosum'. The second column, 'All dictionaries', has a search box containing 'inflammatory bowel disease' and a table with 'id 1: Term B' (13044) and 'inflammatory bowel disease'. The third column, 'Biologica Processes', has a search box containing 'inflammatory bowel disease' and a table with 'id 1: Term C' (219807) and 'lipid metabolism'. The 'Term A' and 'Term B' labels are highlighted with red boxes.

Figure 6: The hypothesis generated by the hypothesis explorer showing that Pyoderma Gangrenosum (PG) may be linked to lipid metabolism.

In the above example (fig 6), we generated a hypothesis. By selecting Pyoderma gangrenosum as 'Term A' and Inflammatory bowel disease' as 'Term B', we search for 'Term C' by selecting biological processes dictionary, 162 terms appeared, and lipid metabolism was a part of these terms. This may mean that lipid metabolism may play a part in pyoderma gangrenosum (PG).

D. Network creation and visualization

Often working with huge numbers of biomedical entities from various dictionaries can be a complex task. On the contrary, IBDDDB provides “Visualize” tab for users to create interactive networks among different entities based on their occurrence in the literature. The user can build the network of their choice by selecting one or more dictionaries at each step and trimming off the links they dislike. The nodes represent the entities from the dictionary and are color-coded while numbering on the edges or links represents the number of PubMed records showing co-occurrence of the entities as highlighted terms. An example of network creation using hypotheses explorer is shown in the Fig 7.

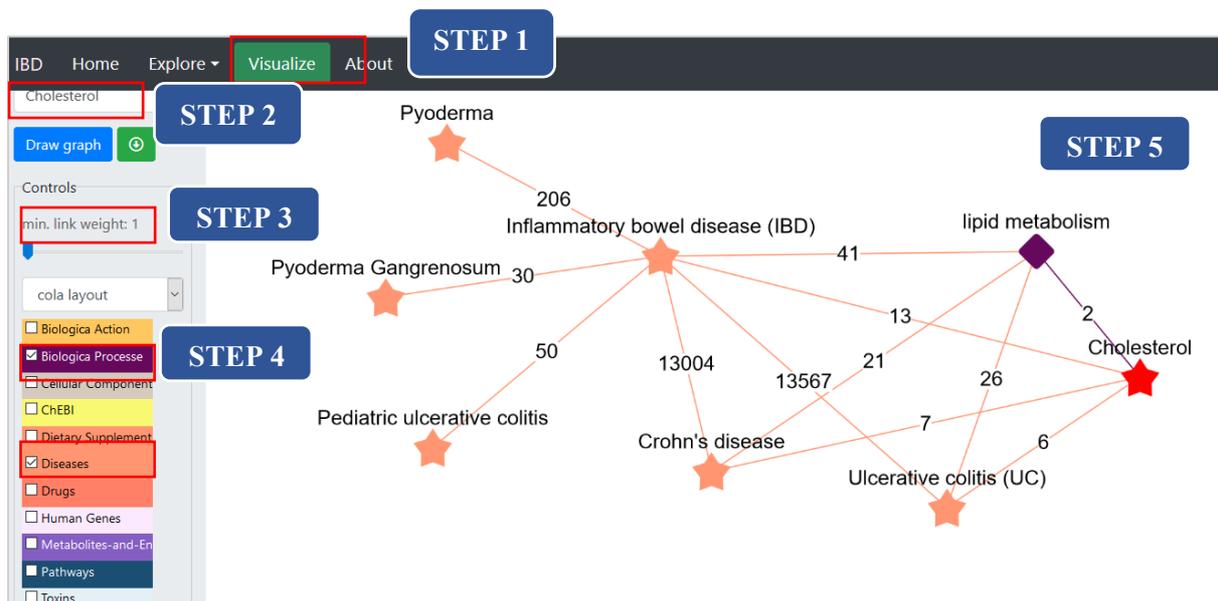


Figure 7: This figure demonstrates the step-by-step network construction to identify the relationship between disease PG and its connection with the pathway lipid metabolism. The red star represents the searched term “cholesterol”. The orange stars represent the “Diseases” dictionary. The purple colored diamond represents the expanded “Biological Processes” dictionary. The colored edges represent the color of their respective dictionary here. The number allocated on each edge showcase the number of publications that link to the associated nodes. The graph can be downloaded by selecting the green box with circled downward arrow to the right of draw graph tab.

E. Downloading Data

This database can make presentations of networks in 6 different layouts and fully exportable data in various formats like. JSON and .PNG as well as printing and direct copying of data to user applications.

The exploration reports like IBD database, biomedical entities, Biomedical entities co-occurrences and hypotheses explorer can be downloaded as Copy, Excel, CSV file, PDF and Print formats.