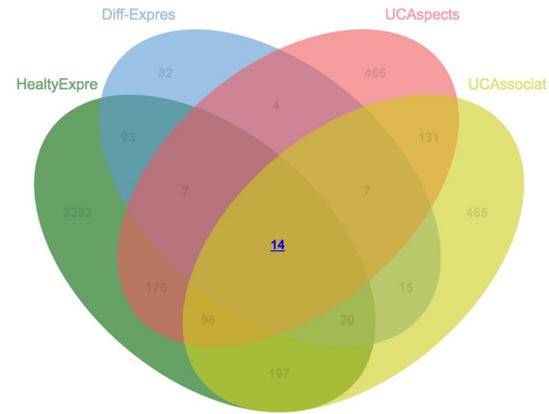


This document has been created as input for the upcoming target evaluation meeting with [REDACTED] where this initial candidate list will be discussed. After the evaluation meeting Euretos will undertake a detailed analysis of selected markers. The following research question is addressed:

*Find potential novel drug targets for Inflammatory Bowel Disease (including FDA drug targets)*

### Approach

The approach leverages the Euretos Knowledge Platform (EKP), where over 175 data sources are consulted simultaneously. The platform generates for each criterion gene lists that meet the criterion. These gene sets are then overlapped to find those gene targets meeting the highest amount of criteria.



### Result

- **4958 potential targets** (i.e. meeting at least one criterion) were analysed
- **Four main groups of criteria were evaluated:**
  1. *Healthy Expression (3997 targets)*: in Colonic Epithelium or Colonic Mucosa
  2. *Differential Expression (242 targets)*: in healthy versus Ulcerative Colitis tissue
  3. *Associations with disease specific characteristics* in particular genes associated with innate immunity, dendritic cells, IL-13 and Primary Sclerosing Cholangitis
  4. *Associations with IBD (947 targets)*: Any gene associated directly in literature with IBD. Please note that this does not include indirect association with the disease e.g. via other concepts. These are found as 'co-occurrence' results.
- **21 targets were found that meet criteria in the first main categories:**
  - 14 targets were found that meet criteria in all four groups
  - 7 targets met all criteria for group 1 to 3, but lacked any direct references to IBD (group 4). These targets are therefore potential novel associations with IBD.

Targets meeting criteria in all 4 groups	Targets without direct IBD association
[REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED]	[REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED]

In the following section these two sets of targets will be detailed. The attached Excel sheet provides this overview including all the gene sets for the various criteria.

The table is divided in two sections. The first section show targets with a direct association with IBD and is sorted on number of publications mentioning the target and IBD. The second section provides targets that meet all main criteria but, at present, lack any direct references to IBD. This second list is ranked on disease association score (see addendum)

Targets		Selection Criteria				Additional Evaluation Criteria				
Target	FDA drug target	Healthy Expression	Differential Expression	Direct relations with UC	Disease characteristics	Pubmed co-occurrence	Gene variants for UC	# of disease relations	# of gene relations	Disease association score
[REDACTED]	No	Colonic Epithelium & Colonic Mucous Membrane	-4.9	Yes	Primary Sclerosing Cholangitis	402	0	619	30	99,05
[REDACTED]	No	Colonic Mucous Membrane	8.0	Yes	IL-13 Interactome	312	1	210	64	98,85
[REDACTED]	Yes	Colonic Epithelium & Colonic Mucous Membrane	9.4	Yes	Dendritic cells, IL-13 interactome	238	1	279	299	98,85
[REDACTED]	Yes	Colonic Epithelium & Colonic Mucous Membrane	5.4	Yes	Primary Sclerosing Cholangitis	193	1	180	18	98,35
[REDACTED]	Yes	Colonic Epithelium	-6.5	Yes	Dendritic cells	80	0	587	288	99,3
[REDACTED]	Yes	Colonic Epithelium	-6.1	Yes	IL-13 Interactome	52	0	276	19	98,35
[REDACTED]	No	Colonic Epithelium	5.07	Yes	Innate Immunity	42	0	1149	66	98,35
[REDACTED]	No	Colonic Epithelium	-6.01	Yes	Dendritic cells	27	0	132	23	98,35
[REDACTED]	No	Colonic Epithelium	-7.38	Yes	innate immunity	23	0	83	29	97,2
[REDACTED]	No	Colonic Epithelium	7.1	Yes	innate Immunity	19	0	7	0	80,22
[REDACTED]	Yes	Colonic Mucous Membrane	10.3	Yes	Primary Sclerosing Cholangitis	19	1	199	9	97,2
[REDACTED]	No	Colonic Epithelium	7.5	Yes	innate Immunity	12	0	1	0	78,72
[REDACTED]	No	Colonic Epithelium	6.8	Yes	innate immunity	10	0	42	15	96
[REDACTED]	No	Colonic Epithelium	-6.4	Yes	innate Immunity	6	0	42	2	92,06

Targets		Selection Criteria				Additional Evaluation Criteria				
Target	FDA drug target	Healthy Expression	Differential Expression	Direct relations with UC	Disease characteristics	Pubmed co-occurrence	Gene variants for UC	# of disease relations	# of gene relations	Disease association score
[REDACTED]	No	Colonic Epithelium	-4.29	No	innate immunity	0	0	0	0	94,61
[REDACTED]	No	Colonic Epithelium	-4.8	No	innate immunity	0	0	216	12	91,51
[REDACTED]	No	Colonic Epithelium	-3.8	No	innate immunity	0	0	52	1	86,71
[REDACTED]	No	Colonic Epithelium	3.4	No	Dendritic cells	3	0	9	3	86,16
[REDACTED]	No	Colonic Epithelium	3.0	No	Dendritic cells	0	0	1	2	84,02
[REDACTED]	No	Colonic Epithelium	-2.7	No	innate immunity	0	0	9	4	80,12
[REDACTED]	No	Colonic Epithelium	7.7	No	innate immunity	0	0	2	2	78,08

Explanation of column headings.

- **Target** - Gene name of the target. See addendum 1 for additional background information,
- **FDA Approved** - If the target is FDA approved (note: not necessarily for IBD)
- **Healthy Expression** - In which healthy tissue the target is expressed: the colonic epithelium and/or the colonic mucous membrane
- **Differential Expression** - The level of differential expression of the target of IBD (in log2fold)

- **Direct relations with IBD** - If there are direct relations known between the target and IBD. Please note that this not include indirect relations between a concept and the disease. These types of relations are found as 'co-occurrence' results (see below)
- **Disease Characteristics** - To which disease associated characteristics the target is associated. The following disease aspects are considered: innate immunity, dendritic cells, IL-13 interaction and Primary Sclerosing Cholangitis
- **Pubmed co-occurrence** - The number of Pubmed abstracts in which the target and IBD co-occur. The targets with direct relations are ranked on this criterion.
- **Gene variant for IBD** - Whether the are known gene variants associated with IBD
- **# of disease relations** - The number of other diseases the target is associated with
- **# of gene relations** - The number of other genes the target is interacting with
- **Disease association score** - The target - disease association score (see addendum 2). The targets without direct relations are ranked in this criterion.

## Next Steps

These targets will be discussed during the target survey meeting. In preparation to this meeting, the Euretos Gene Disease Analysis application provides detailed information on each specific gene in the context of IBD. During the targets survey meeting the model will be refined based on expert feedback after which the key targets will be selected for further, more detailed analysis.

## Addendum - Supporting Background

### 1 - Target Classifications

Below are additional classifications of each target. They are presented in the same order as in the main table.

Target	Classification	FDA Drug Target
[REDACTED]	Defense/immunity protein Predicted membrane proteins Receptor Predicted secreted proteins Receptors, cytokine Cancer-related genes Candidate cardiovascular disease genes Immunoglobulin receptor superfamily	No
[REDACTED]	Chemokine Plasma proteins Signaling molecule Cytokine Predicted secreted proteins Cancer-related genes Candidate cardiovascular disease genes	No
[REDACTED]	Interleukin superfamily Plasma proteins Signaling molecule Cytokine Cancer-related genes Candidate cardiovascular disease genes	Yes
[REDACTED]	Plasma proteins Cancer-related genes	Yes

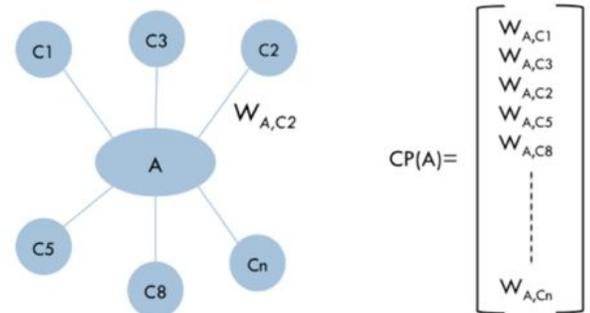
[REDACTED]	Lyase Plasma proteins Cancer-related genes Candidate cardiovascular disease genes Hydro-lyases	Yes
[REDACTED]	Plasma proteins Cytoskeleton related proteins Signaling molecule Intracellular calcium-sensing protein Cancer-related genes Calmodulin	No
[REDACTED]	Plasma proteins Cytoskeleton related proteins Signaling molecule Intracellular calcium-sensing protein Cancer-related genes Calmodulin	No
<b>Target</b>	<b>Classification</b>	<b>FDA Drug Target</b>
[REDACTED]	Plasma proteins Predicted secreted proteins Cancer-related genes Candidate cardiovascular disease genes	Yes
[REDACTED]	Small gtpase RAS pathway related proteins Cancer-related genes Gtp-binding proteins Enzyme modulator	No
[REDACTED]	Cell adhesion protein Cancer-related genes	No
[REDACTED]	Plasma proteins Transfer/carrier protein Predicted secreted proteins Cancer-related genes Candidate cardiovascular disease genes Isomerase	No
[REDACTED]	Plasma proteins Transporters Cancer-related genes	Yes
[REDACTED]	Plasma proteins Predicted secreted proteins	No
[REDACTED]	Plasma proteins Predicted secreted proteins Cancer-related genes Candidate cardiovascular disease genes	Yes
[REDACTED]	Plasma proteins Predicted secreted proteins Transporters	No
[REDACTED]	Predicted membrane proteins Free fatty acid receptor 3 G-protein coupled receptors Receptor	No
[REDACTED]	Hydrolase Oxidoreductase Receptor Oxidase Predicted secreted proteins Cancer-related genes Endopeptidases	No
[REDACTED]	Transcription factors Cancer-related genes Transcription factor	No
[REDACTED]	Plasma proteins Predicted membrane proteins Cytoskeleton related proteins Protein kinase Transferase Transfer/carrier protein Intracellular calcium-sensing protein Non-receptor serine/threonine protein kinase Annexins Phosphotransferases Calmodulin	No
[REDACTED]	Signaling molecule RAS pathway related proteins Cancer-related genes Growth factor	No

[REDACTED]	Membrane transport proteins Defense/immunity protein Plasma proteins Transfer/carrier protein Predicted secreted proteins Cancer-related genes Candidate cardiovascular disease genes Apolipoproteins	No
[REDACTED]	Predicted secreted proteins	No
[REDACTED]	Nuclear hormone receptors Transcription factors Nucleic acid binding Receptor Transcription factor Nuclear receptors	Yes
[REDACTED]	Ligase Cancer-related genes Ubiquitin-protein ligase	No
Target	Classification	FDA Drug Target
[REDACTED]	Plasma proteins Predicted membrane proteins Nucleic acid binding Cytoskeleton related proteins Hydrolase Chromatin/chromatin-binding protein Cancer-related genes Kinase modulator Dna binding protein Enzyme modulator	No
[REDACTED]	Plasma proteins Predicted membrane proteins Cancer-related genes Kinase modulator Enzyme modulator	No
[REDACTED]	Hydrolase	No
[REDACTED]	Cancer-related genes	No
[REDACTED]	Nucleotidyltransferase Defense/immunity protein Nucleic acid binding Transferase	No

## 2 - Disease association score

An objective measure comparing the number and specificity of indirect relations between two (*related or unrelated*) biological concepts against a reference set of related concepts of the same type. The resulting score is the is percentile rank against the frequency distribution of a reference set.

The Disease Association Score ranks two concepts (*whether they have a direct relation or not*) in terms of their indirect relations, against a reference set of concepts of the same semantic category for which direct relations do exist. It provides an objective measure of the level of mutual information two concepts share. In cases where no direct relation exists the Disease Association Score act as a predictive statistical function as it compares the level of mutual information of both existing and non existing relations.



The Disease Association Score uses the methodology of “concept profile analysis” as developed at the Leiden University Medical Centre (LUMC), and described in scientific publications (eg van Haagen HH et al ; “ Novel protein-protein interactions inferred from literature context” , 2009).

The score is calculated as follows:

- **Selected Concepts:** The two selected biological concepts for which the functional association will be calculated. Each of the concepts belongs to a **semantic category**.
- **Reference set:** set of existing relations in the dataset where the concepts in the relationship have the same **semantic category** assignment as the Selected Concepts that are being compared.
- **Weight:** distance score between 2 related concepts, indicate similarity (represented by the Jaccard index) of the connected concepts based on context in the graph environment.
- **Concept profile:** Summation of all weights associated to a single concept
- **Concept profile analysis score:** the inner product of the weights associated with relations that the Selected Concept have in common.
- **Functional Association Score:** the percentile rank of the concept profile analysis score on the 2 concepts in the hypothesis, compared to the frequency distribution of the concept profile analysis scores of the reference set.

