Date: [redacted]
Version: 0.2

This initial indications survey assesses is part of the research project between Euretos and [redacted]. It addresses the following research question:

Find potential indications for the [redacted] target

Result

Using the Euretos Knowledge Platform (EKP), where over 175 data sources are consulted simultaneously, 97 potential indications were found and assessed. Of these, 15 indications are ranked highest as either differential expression or gene variant associations with the [redacted] target were found. These are outlined below (see approach section for scoring model):

	INDICATION		DIFFERENTIAL EXPRESSION	GENETIC VARIANTS	LITERATURE		TOTAL SCORE	# of Publications
	Disorder	Disease categories			Publications	association score		
1	asthma	Diseases of the respiratory system		2	0,84	0,96	3,80	172
2	chronic obstructive airway disease	Diseases of the respiratory system	2		0,48	0,96	3,44	100
3	carcinoma of lung	Neoplasms	2		0,66	0,48	3,14	135
4	non-small cell lung carcinoma	Neoplasms	2		0,43	0,69	3,12	88
5	rheumatoid arthritis	Diseases of the musculoskeletal system and connective tissue		2	0,15	0,96	3,11	32
6	rhinitis	Finding		2	0,20	0,91	3,11	42
7	bronchial hyperreactivity	Finding		2	0,12	0,92	3,04	26
8	lupus erythematosus, systemic	Diseases of the musculoskeletal system and connective tissue		2	0,08	0,95	3,03	16
9	Sarcoidosis	Diseases of the blood		2	0,09	0,94	3,03	18
10	iga glomerulonephritis	Finding		2	0,12	0,91	3,03	25
11	chronic sinusitis, nos	Diseases of the respiratory system	2		0,06	0,91	2,97	12
12	allergic disposition	Finding		2	0,15	0,74	2,88	31
13	bronchiolitis, viral	Finding		2	0,00	0,80	2,80	0
14	dermatitis and eczema	Finding		2	0,00	0,73	2,73	0
15	cancer of head and neck	Neoplasms	2		0,02	0,36	2,38	5

The full list of 97 indications is available in a separate document. For the top 15 indications a full, in depth overview is provided separately including all available references to:

- Differential expression experiments for the [redacted] target and the indication
- Gene variants of the [redacted] target for indications in literature
- All publications mentioning the [redacted] target and the top 15 disease indications.

These references are provided as clickable links to the supporting evidence in the separate detailed document.

Overview of the [redacted] target

Gene name	[redacted]
Synonyms (selection)	[redacted], [redacted], [redacted], [redacted], [redacted]
Protein names (selection)	[redacted], [redacted] [redacted] [redacted], [redacted], [redacted]
Classification	cytokine
Expression	Predominantly in [redacted]

Differential expression summary of the [redacted] target

Experiment Type	Characteristic	Differential Expression	Detail
compound	Decitabine	overexpression	Epithelial acinar cell
compound	Lipopolysaccharides	overexpression	Alveolar macrophage
disease	NSCLC	downregulated	Multiple experiments
disease	Lung carcinoma	donwregulated	Multiple experiments
disease	Interstitial lung disease	overexpression	Lung tissue
disease	Glioma	downregulated	human embryonic stem cell-derived neural progenitor cell
disease	Breast cancer	downregulated	cn
disease progression	Prostate cancer	downregulated	Benign -> primary cancer
disease	COPD	downregulated	Alveolar macrophage
disease	Interstitial Cystitis	overexpression	Bladder
disease	Cancer of nasopharynx	downregulated	Nasopharynx epithelial cell

Approach

The initial indications survey was created using the below methodology. The starting point are all known *disease annotations* associating the target with a potential indication. For these initial indications the following aspects will be analysed:

1. *Differential expression* results for the disease indication (disease tissue vs healthy) where the selected target is either down-regulated or up-regulated with a log2fold change of at least 1.
2. *Gene variant findings* where the selected target is mentioned as a variant associated with the disease indication
3. The number of *co-occurrence literature references* mentioning the target and the indication. This will also include all synonyms of the target and the disease indication.
4. The *disease association score* which provides an objective measure of target - indication interaction based on indirect associations (see addendum 1)

Each indication as ranked on these criteria using the following model:

Evaluation criteria	Score
1 - Differential expression: are there experiments known where the target is up or down regulated when comparing healthy to diseased tissue. If so, a score of 2 is applied.	0 or 2
2 - Gene variants: are there known gene variants for the disease? If so, a score of 2 is applied.	0 or 2
3A - Literature - Direct co-occurrence: In how many Pubmed abstracts are the gene and the disease mentioned together (i.e. directly co-occur). The score is calculated as the relative number of publications compared to the top result in the list of indications.	Between 0 and 1
3B - Literature - Indirect association score: provides an objective ranking between 0 and 1 for the level of indirect relations a gene and disease have (see Addendum 1)	Between 0 and 1

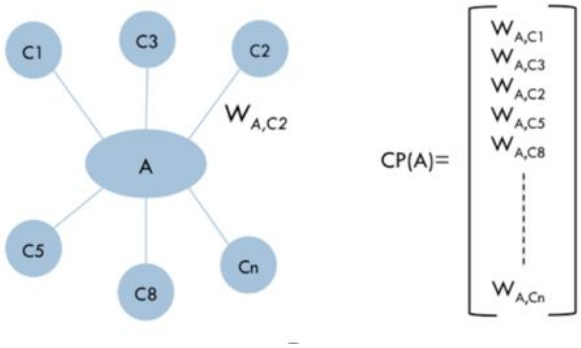
Next Steps

The full list of 97 indications is available in a separate document. For the top 15 indications a full, in depth overview is provided separately including all available references. These indications will be discussed during the target survey meeting. If the references for other disease indications are required they can be delivered on demand. The Gene Disease Analysis application is available providing in depth capabilities to assess the [redacted] target involvement for each indication if required.

Addendum - Supporting Background

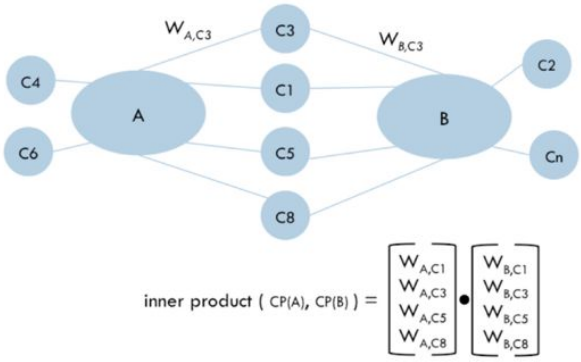
1 - Disease association score

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.