Knowledge Acquisition and Application for biological Impact Analyses in an Industrial Setup

Sam Ansari
Philip Morris International R&D, Philip Morris Products S.A.
(part of Philip Morris International group of companies)

28 April 2016
eKNOW 2016 – Industry Research

Philip Morris International is the sole source of funding and sponsor of this project.

Overview

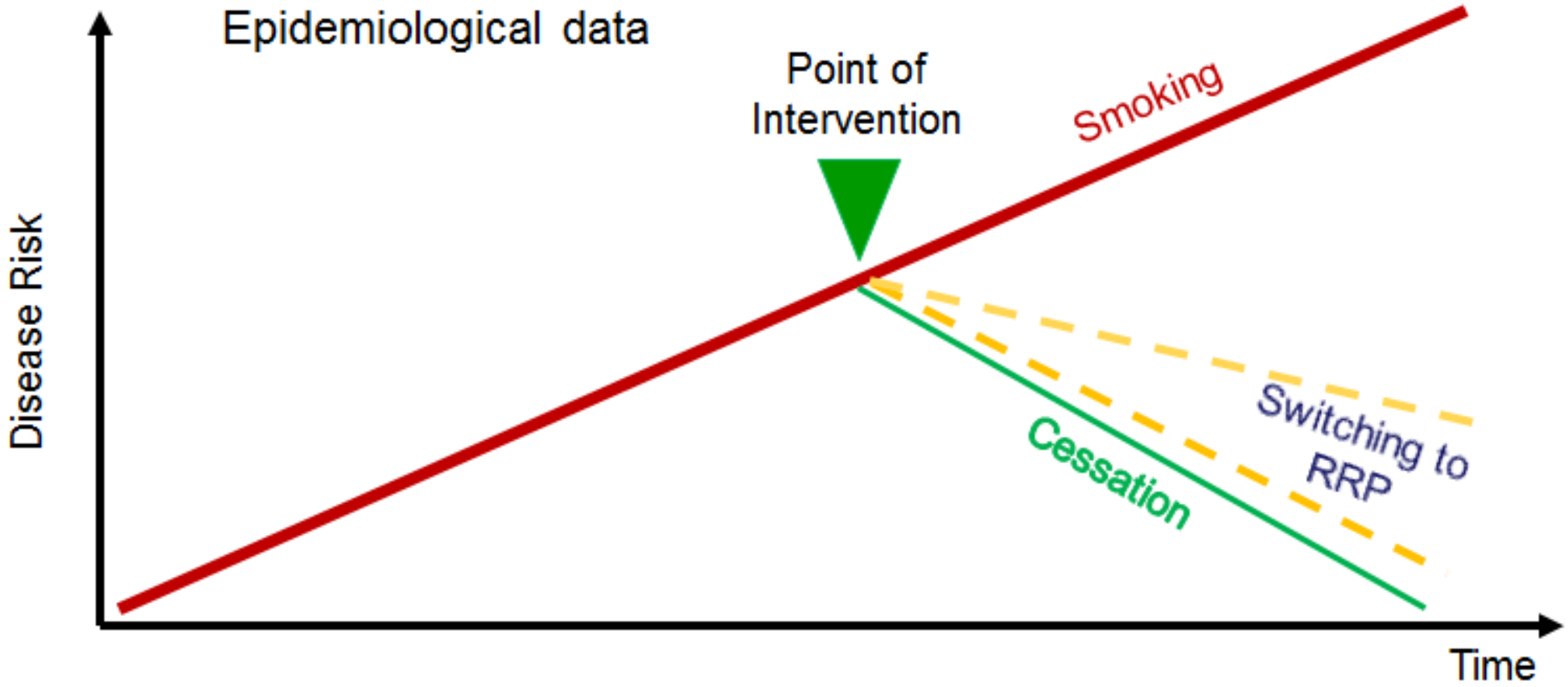
Background – PMI R&D

Smoking causes serious diseases such as cardiovascular diseases, lung cancer, and chronic obstructive pulmonary disease.

Philip Morris International is therefore developing novel products that may have the potential to reduce individual risk and population harm in comparison to smoking cigarettes.

To determine whether such products have the potential to reduce disease risk, we compare their biological impact with that of a standard reference cigarette (3R4F) on a mechanism-by-mechanism basis.

Background – Smoking Cessation as Benchmark



Note: Reduced-Risk Products ("RRPs") is the term we use to refer to products that have the potential to reduce individual risk and population harm in comparison to smoking combustible cigarettes

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Systems Toxicology

www.pmiscience.com

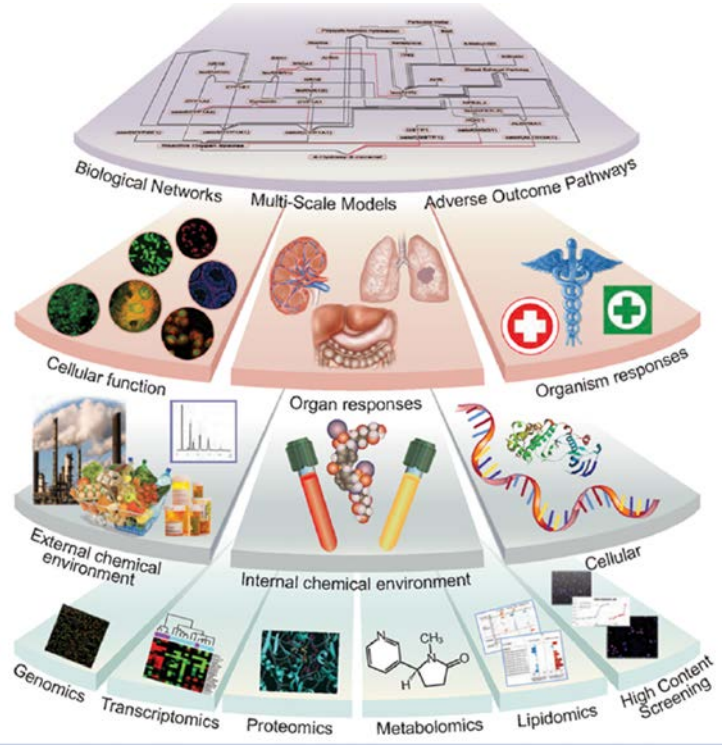
- considers biological systems as a whole and aims at elucidating detailed **biological mechanisms** that link exposure to active substances with their adverse consequences
- integrates classic toxicology approaches with the quantitative analysis of molecular and functional changes
- combines high throughout methods with advanced computational methods
- enables the shift to a new paradigm for risk assessment which is the future tox or 21st century toxicology (*Product Assessment*)

Computational models

Apical measurements

Molecular measurements

Enabling technologies

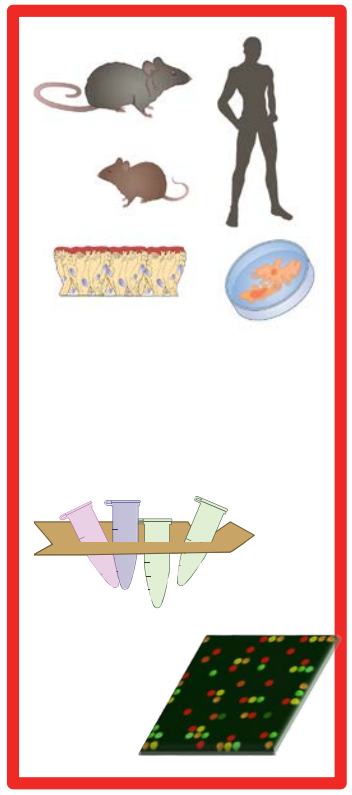


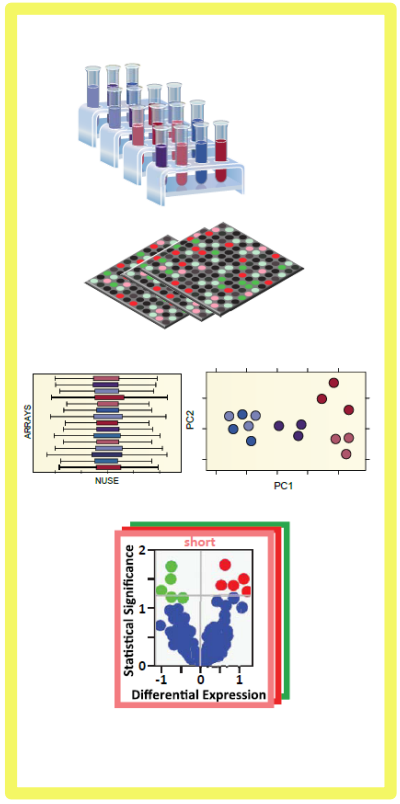
Sturla et al., 2014; Chemical research in toxicology 27: 314-329
Hartung et al., 2012; ALTEX. 29(2):119-28

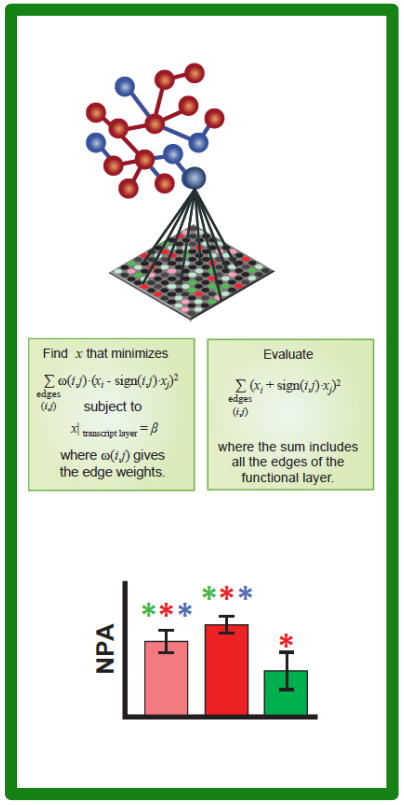
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

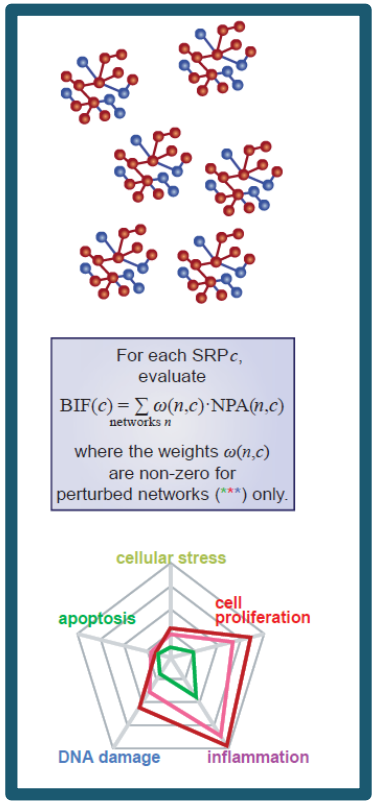
Systems Toxicology – Assessment Approach

Experimental data production → **Compute systems response profiles** → **Identify perturbed biological networks** → **Compute network perturbation amplitudes** → **Compute product biological impact**







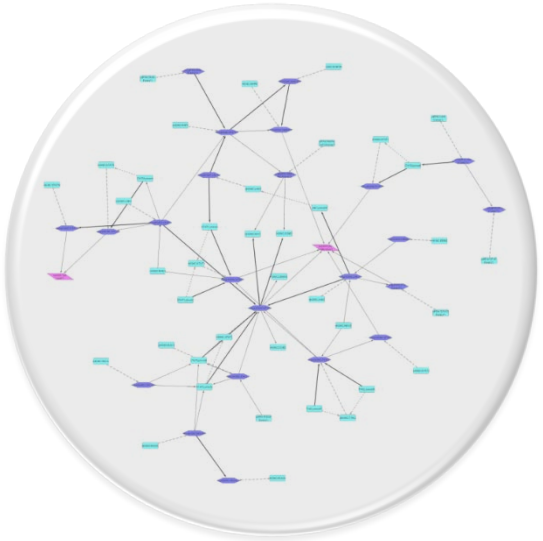


Hoeng et al., 2012; Drug discovery today 17: 413-418

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

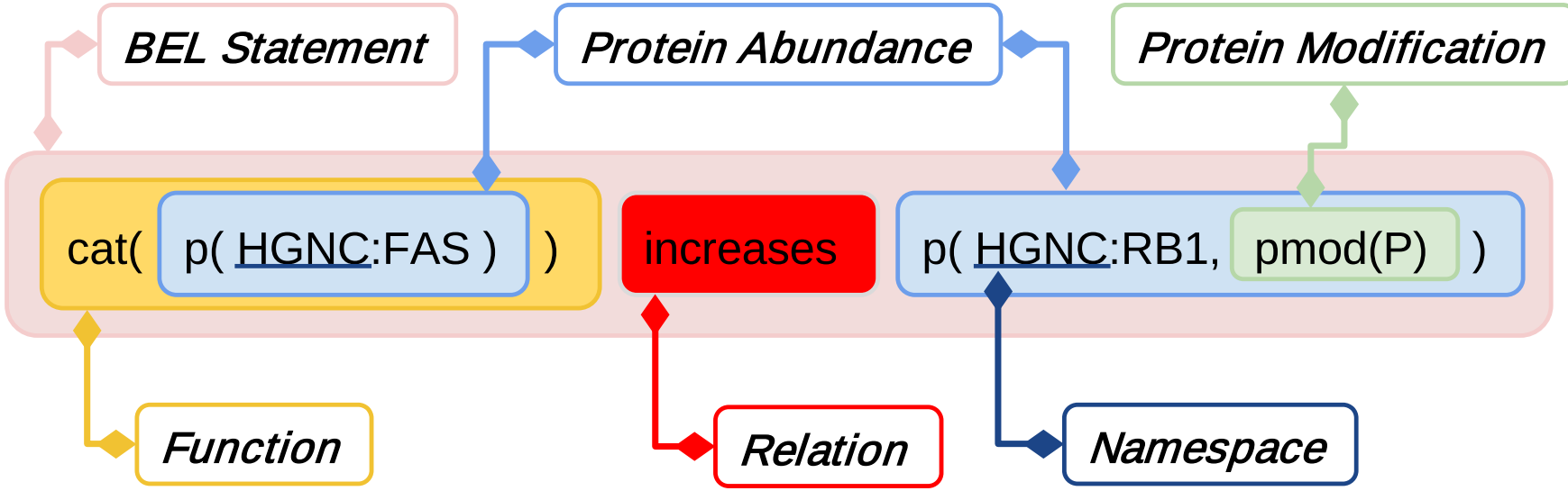
Knowledge Networks

From Literature to Computable Knowledge



Knowledge Networks - Biological Expression Language (BEL)

BEL Statement



Citation

```
SET Citation = {"PubMed", "Regulation of Rb and E2F by signal transduction cascades:
divergent effects of JNK1 and p38 kinases." , "EMBO J. 1999 Mar 15;18(6):1559-70." ,
"10075927"}
```

Support

```
SET Evidence =
        "Fas stimulation of Jurkat cells is known to induce p38 kinase and we find
         a pronounced increase in Rb phosphorylation within 30 min of Fas stimulation"
```

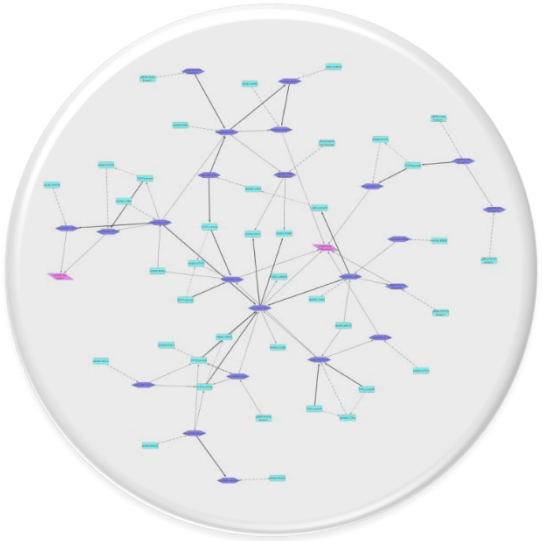
Experiment Context

```
SET Tissue = "jurkat cells"
```

Slater, T., 2014; Drug Discov. Today, 19, 193–198

Knowledge Networks

From Literature to Computable Knowledge



Knowledge Extraction – Overview

Manual and Text Mining | Manual | BELIEF | OpenBEL framework | sbvIMRPOVER NVC

1 Selection of relevant articles (Triage)

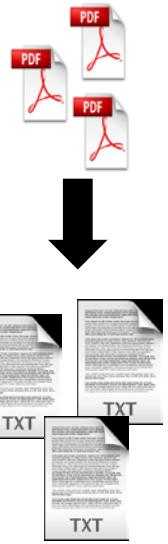
2 Articles Pre-Processing

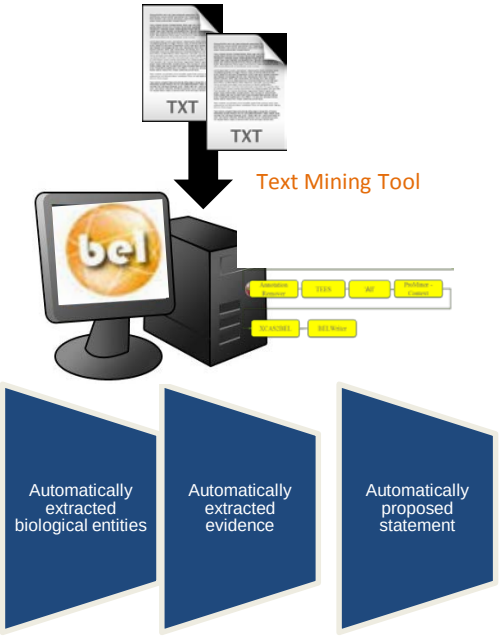
3 Text Mining Pipeline

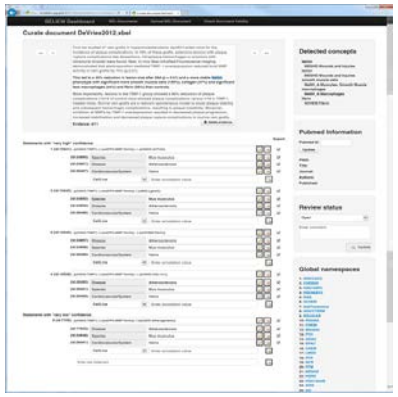
4 Curation Interface

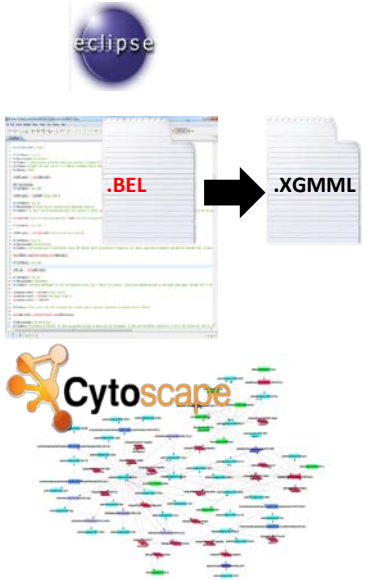
5 Knowledge Network Model Building

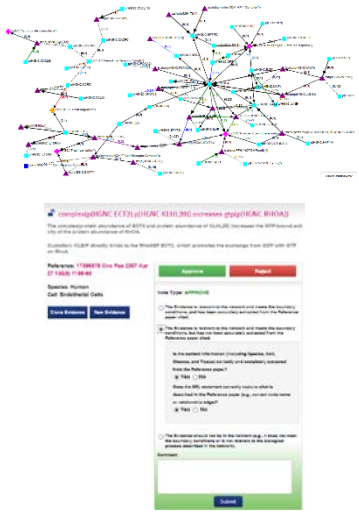
6 Network Model Verification







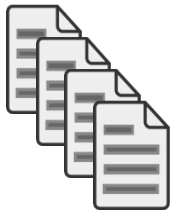




PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Knowledge Extraction - BELIEF Pipeline

Expectation



Full Text

NER & Normalization

Relation Extraction

Implementation BELIEF Pipeline

TXT

Various NLP tools

Unstructured Information Management Architecture
An Apache Project

NER with multiple dictionaries

RE pre-process

Relation extraction tools

BEL Writer

BEL

Szostak, J. et al., 2015; Database (Oxford)
Madan, S. et al., 2016; Database (Oxford)

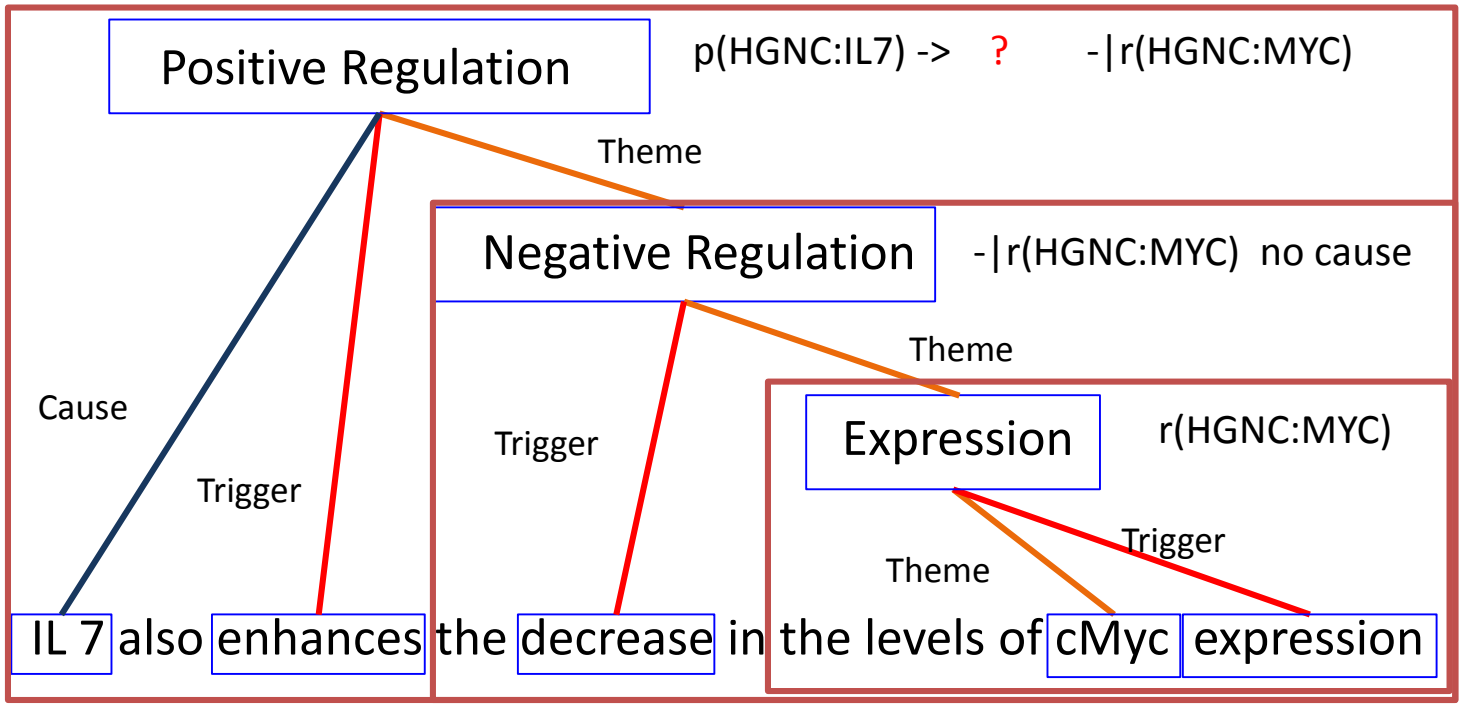
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Knowledge Extraction - Named Entity Recognition

Entity class	Resources
Human genes/proteins	EntrezGene/Uniprot
Mouse genes/proteins	EntrezGene/Uniprot
Rat genes/proteins	EntrezGene/Uniprot
Protein family names	OpenBEL
Protein complex names	OpenBEL
Protein complex names	Gene Ontology
Biological processes	Gene Ontology
Chemical names	OpenBEL
Chemical names	ChEBI
Chemical names	ChEMBL
Disease names	MeSH
Anatomical names	MeSH
Cell lines	Cell Line Ontology
Cell structures	MeSH

Knowledge Extraction – Relation Extraction

- The BioNLP shared tasks delivers a very detailed annotation for relationship extraction similar to the information needed for BEL (TEES2.1):



- Simpler binary classification (LibLINEAR):

IL 7 also enhances the decrease in the levels of cMyc ….

IL7 - cMyc Relation: Yes

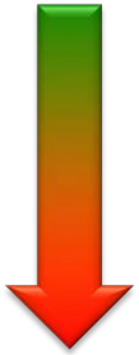
p(HGNC:IL7) - - p(HGNC:MYC)

Classifies if a relation between 2 entities is existing but gives no information about the direction or type

Björne, J. et al., 2012; BMC Bioinformatics, 13 Suppl 1, S4
Fan, R. E. et al., 2008; J. Mach. Learn., 9, 1871–1874

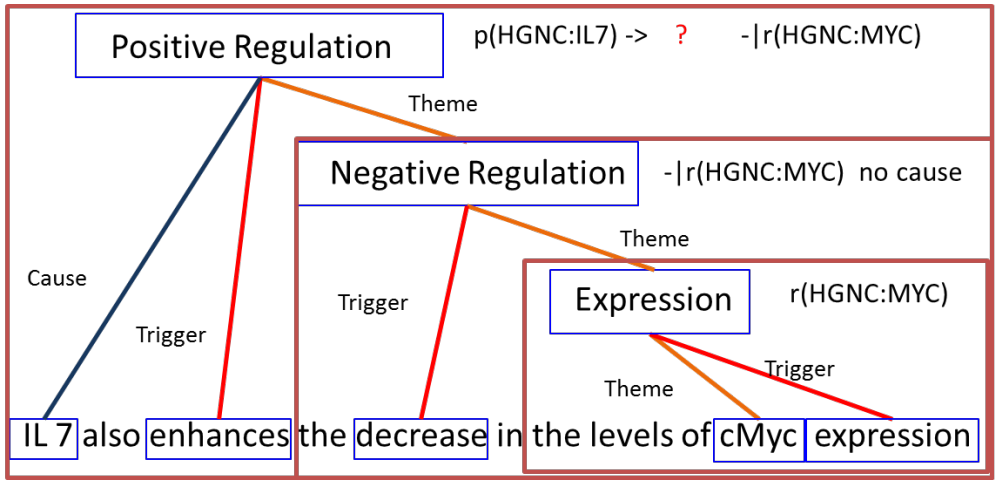
BELIEF Pipeline - Limitations

- NLP (Sentence Detection ~6% error)
- NLP (Tokenization ~8% error)
- NER (Different Classes ~15% error)
- RelationExtraction (Multi-step ~25% error)



complexity

Propagated error



Class	Precision	Recall	F-score
Term	81.34	72.67	76.76
Function	51.16	33.33	40.37
Relationship	67.37	31.68	43.10
Statement	59.15	20.79	30.77

Generated Statement:

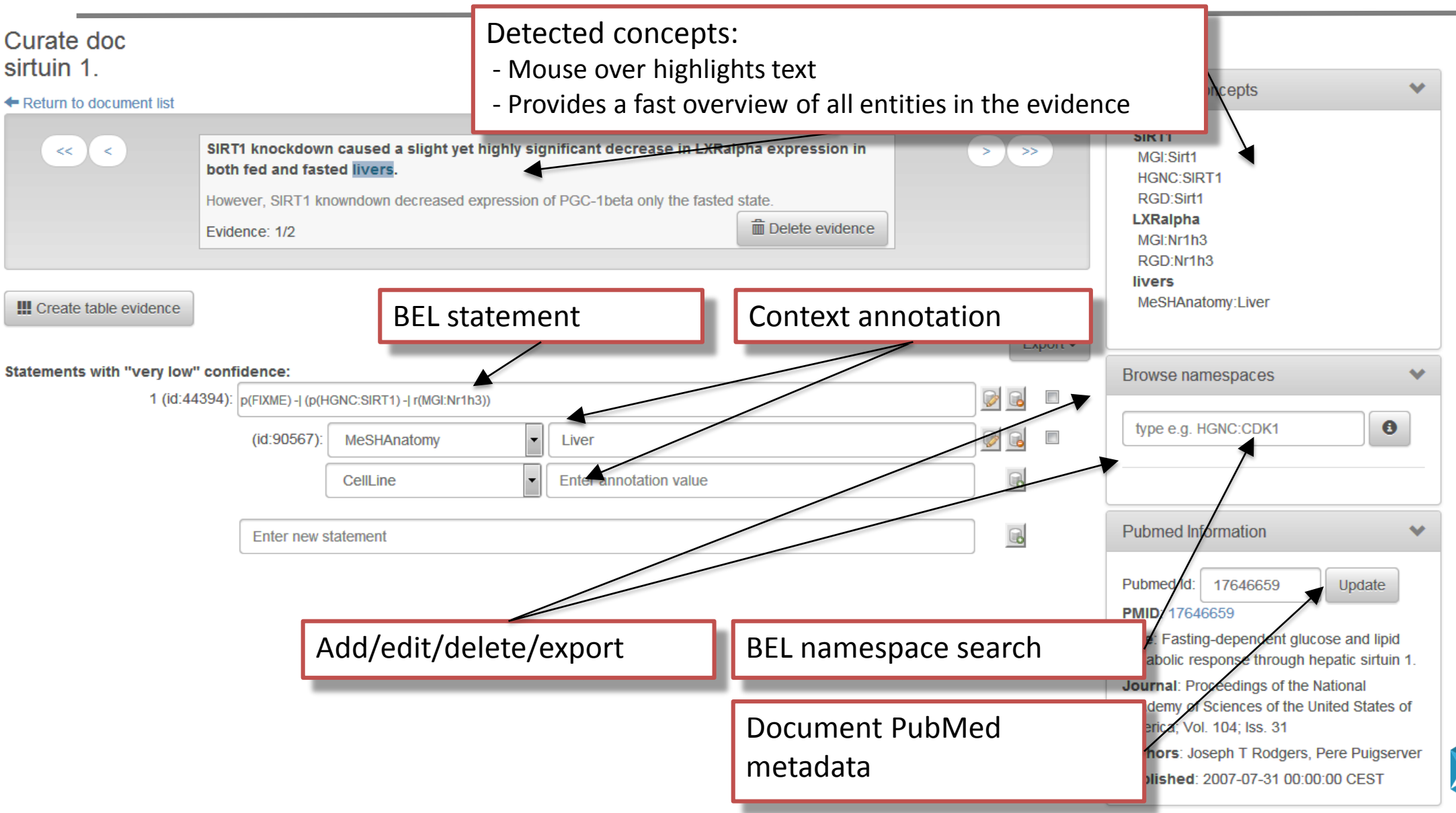
p(HGNC:IL7) -> p(Placeholder) -|r(HGNC:MYC)

Manual Statement:

p(HGNC:IL7) -| r(HGNC:MYC)

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Knowledge Extraction – Curation Interface



Detected concepts:
- Mouse over highlights text
- Provides a fast overview of all entities in the evidence

BEL statement

Context annotation

Add/edit/delete/export

BEL namespace search

Document PubMed metadata

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

http://belief.scai.fraunhofer.de/BeliefDashboard/

Knowledge Extraction – Curation Interface

BELIEW Dashboard | Home | BEL-Documents | Upload BEL-Document | Projects | Documents

Create Project

Name

Enter name

Description

Enter description

Create

BELIEW Dashboard | Home | BEL-Documents | Upload BEL-Document | Projects | Documents

- List
- Create

Document Lists

Search:

Id	Pubmed Id
1	2234234
2	12165281
Id	Pubmed Id

Showing 1 to 2 of 2 entries

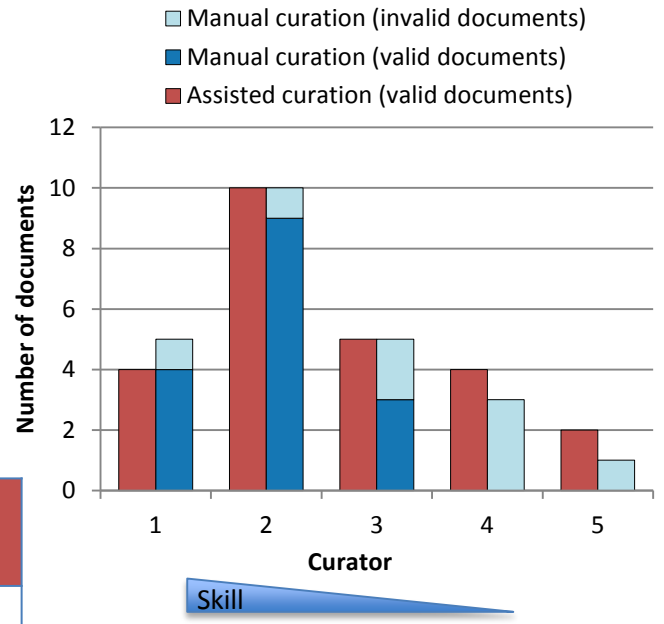
Show 10 entries Search:

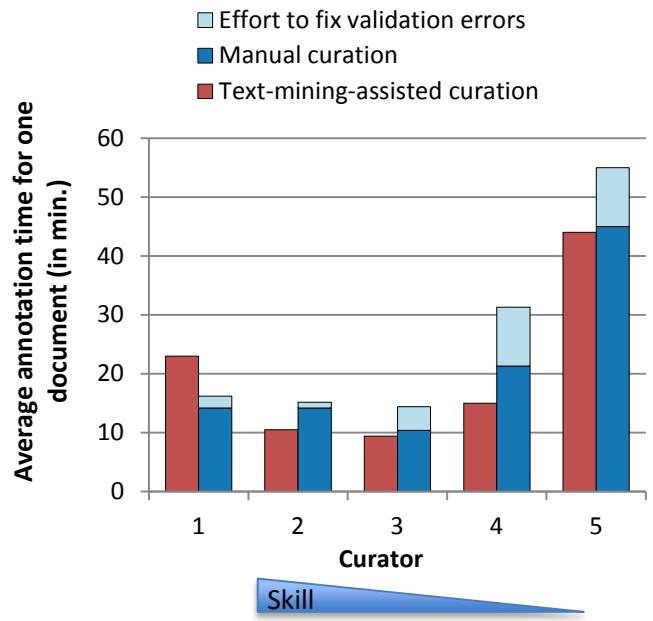
Id	Statement	EvidenceId	Evidence	Export	Curate	Delete
14580	p(HGNC:PARK7, sub(L,166,P)) -> path(MESHD:"Parkinsonian Disorders")	1	In the present study, we investigated the effects of edaravone on the neurotoxicity in PD-induced isoforms of DJ-1 containing the mutation L166P.	☑	✎	✖
14581	p(HGNC:PARK7) -- p(MESHD:"Parkinsonian Disorders")	1	In the present study, we investigated the effects of edaravone on the neurotoxicity in PD-induced isoforms of DJ-1 containing the mutation L166P.	☐	✎	✖
14582	a(CHEBI:edaravone) -> p(HGNC:SLC18A2)	2	Interestingly, our result also demonstrated that edaravone was able to up-regulate VMAT2 expression in N2a cells in a dose-dependent manner.	☑	✎	✖
14583	a(CHEBI:edaravone) -| bp(MESHPP:"Oxidative Stress")	2	Interestingly, our result also demonstrated that edaravone was able to up-regulate VMAT2 expression in N2a cells in a dose-dependent manner.	☑	✎	✖
14584	a(CHEBI:edaravone) -| path(MESHD:"Parkinsonian Disorders")	3	Our findings enhance the understanding of the neuro-protective effects of edaravone in cell models and suggest that edaravone offers significant protection in a PD-related in vitro model.	☑	✎	✖
14586	a(CHEBI:edaravone) -| bp(MESHPP:Apoptosis)	2	Interestingly, our result also demonstrated that edaravone was able to up-regulate VMAT2 expression in N2a cells in a dose-dependent manner.	☑	✎	✖
Id	Statement	EvidenceId	Evidence	Export	Curate	Delete

Showing 1 to 6 of 6 entries

Previous 1 Next

BELIEF Dashboard – Performance and Impression





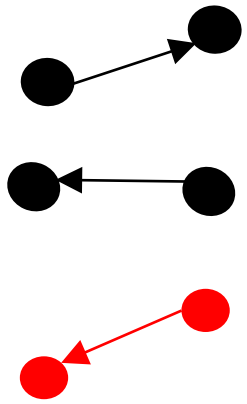
Knowledge Type detected	Manual Curation Network Knowledge	Assisted Curation Network Knowledge
Number of nodes	63	76
Number of edges	94	128
Number of evidences	21	26
Chemical abundance	5	7
Protein abundance	30	32
RNA abundance	2	3
Complex abundance	3	5
Biological process	10	14

Fluck, J., Madan, S., Ansari, S., et al. (2016) Submitt. to Database J. Biol. Databases curation, 2016, Training corpora for the extraction of causal relationships coded in Biological Expression Language (BEL).

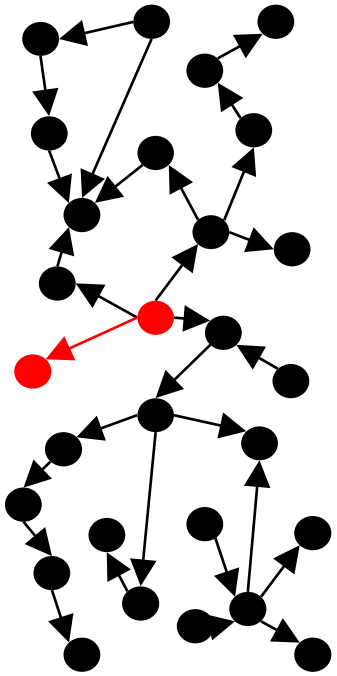
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

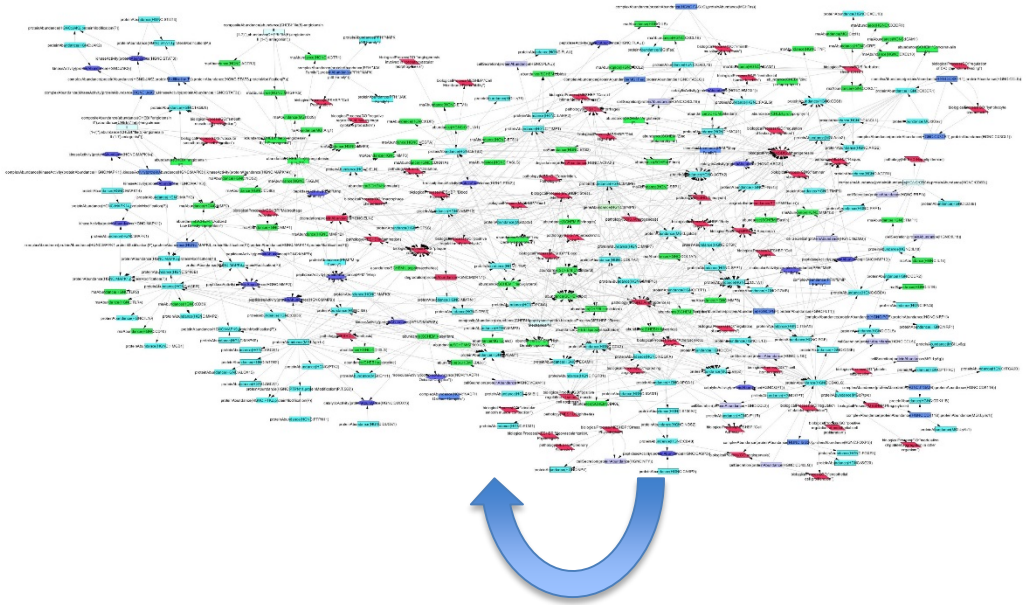
http://openbel.org/

Network Building



p(HGNC:IL7) -| r(HGNC:MYC)





Gap Identification and Trimming

Publication

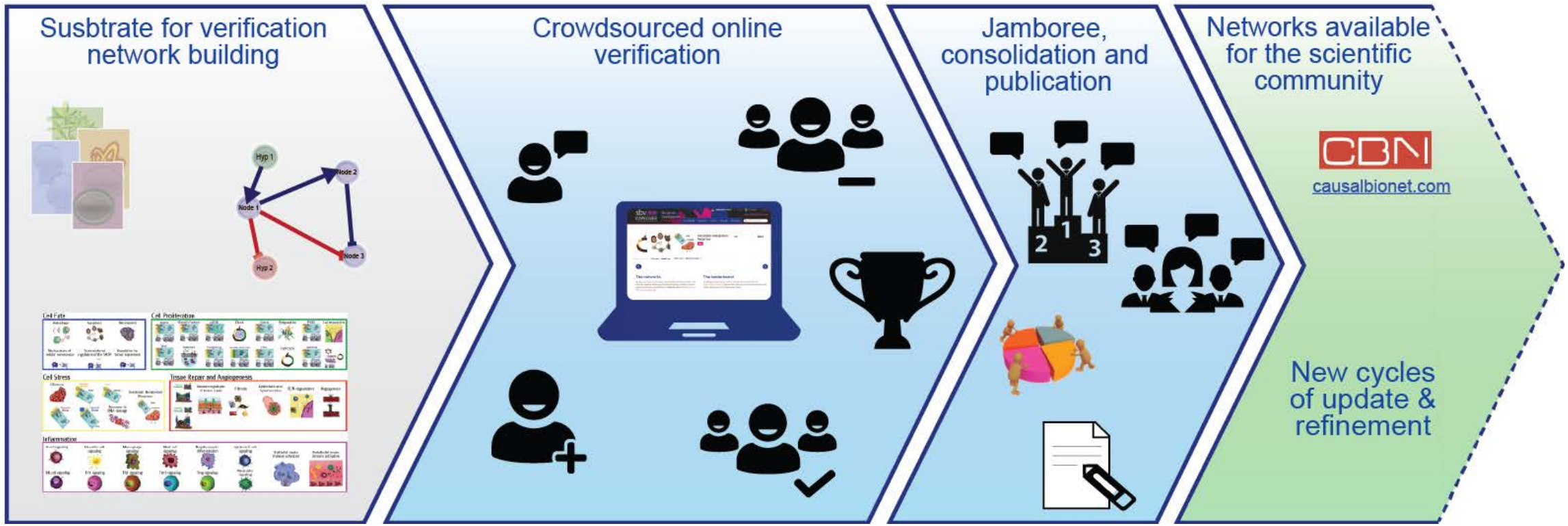
OpenBEL framework

Cytoscape

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

https://sbvimprover.com/challenge-3/challenge

Crowd Verification - Overview

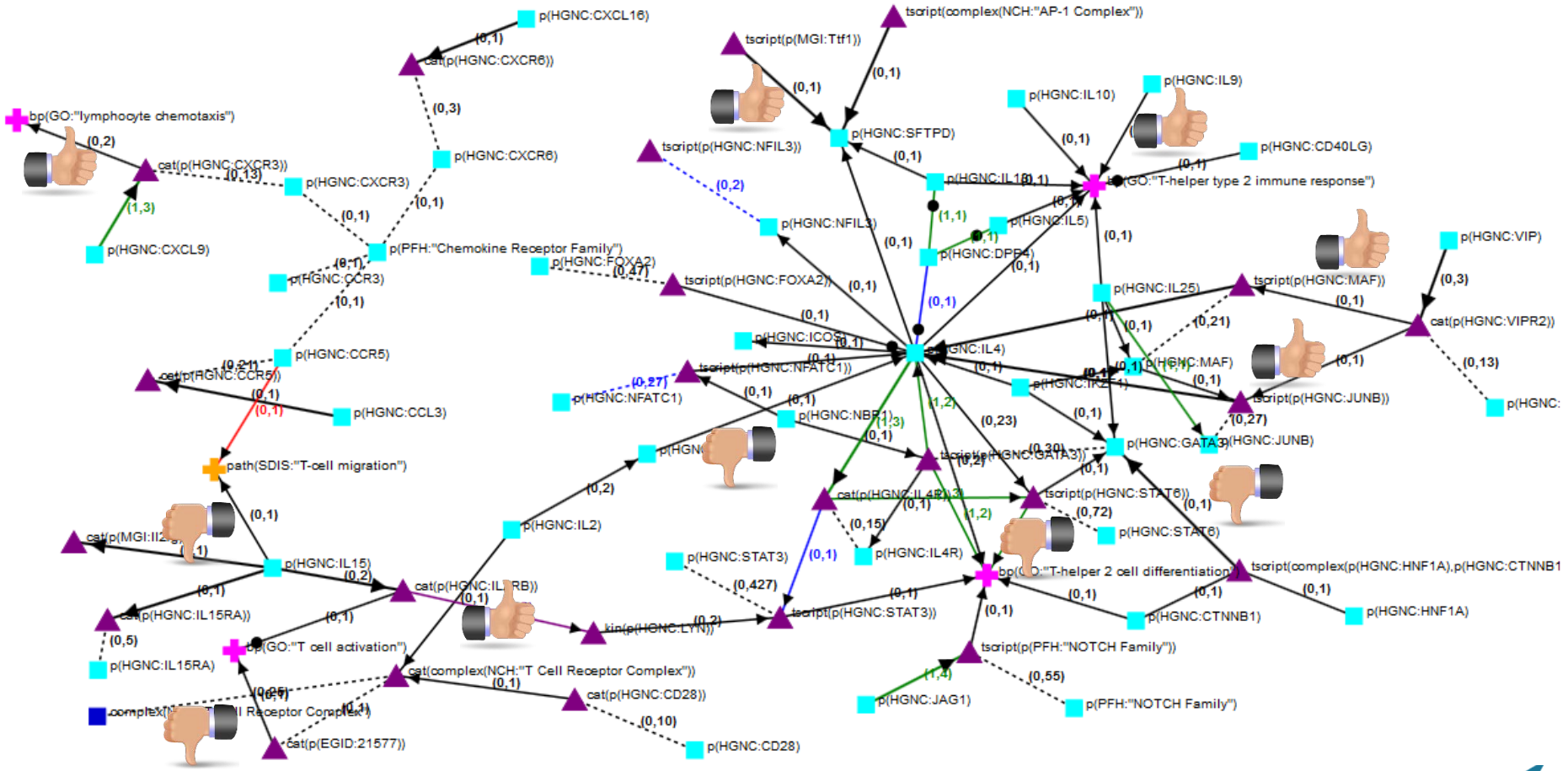


Boue, S., Fields, B., Hoeng, J., et al. (2015), F1000Res. 2015 Jan 29;4:32, Enhancement of COPD biological networks using a web-based collaboration interface.

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Network Verification - Concept

Vote on evidence, create new edges, add missing nodes

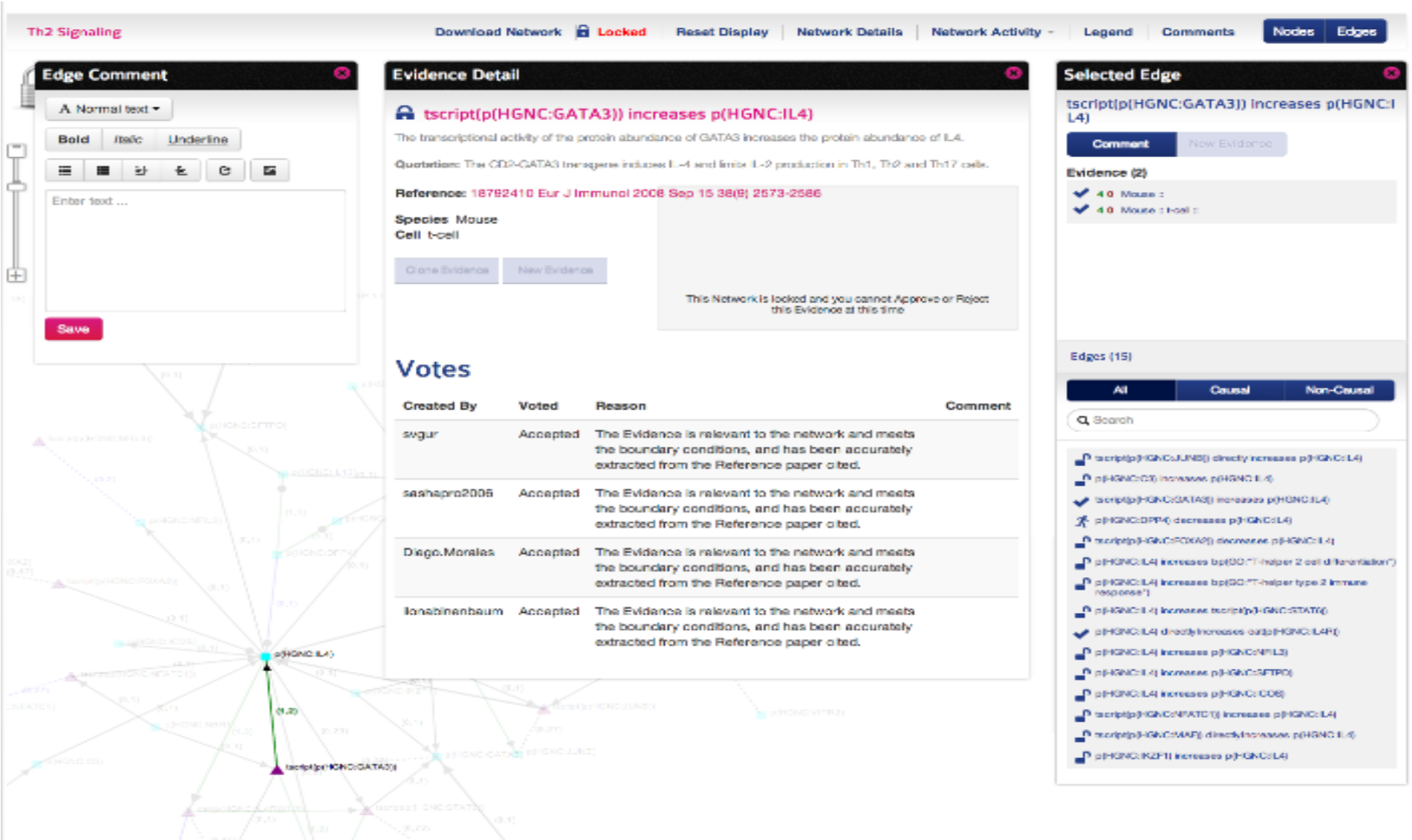


Network Verification – Selection of Sub-network

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

https://sbvimprover.com/challenge-3/challenge

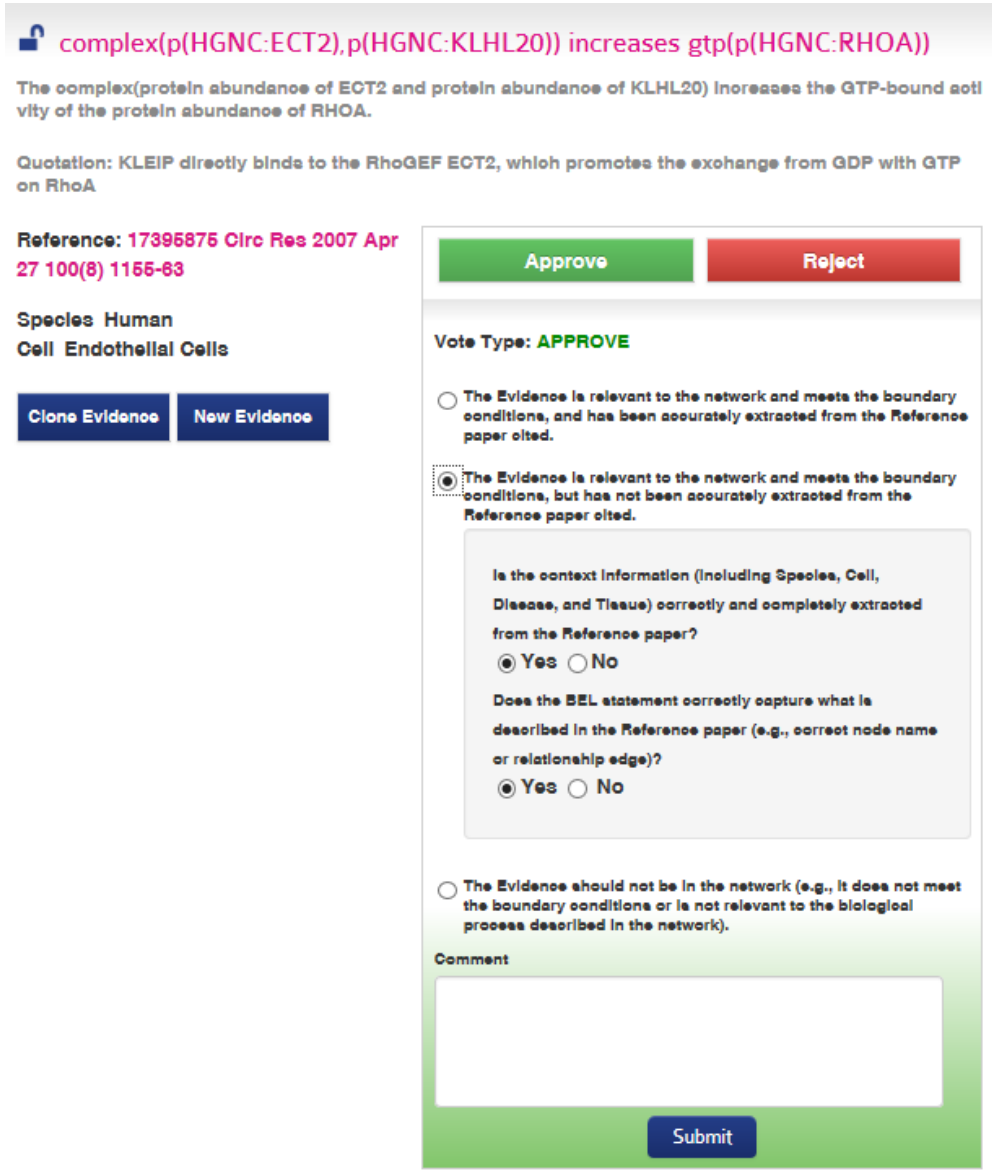
Network Verification – Review of Triples



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

https://sbvimprover.com/challenge-3/challenge

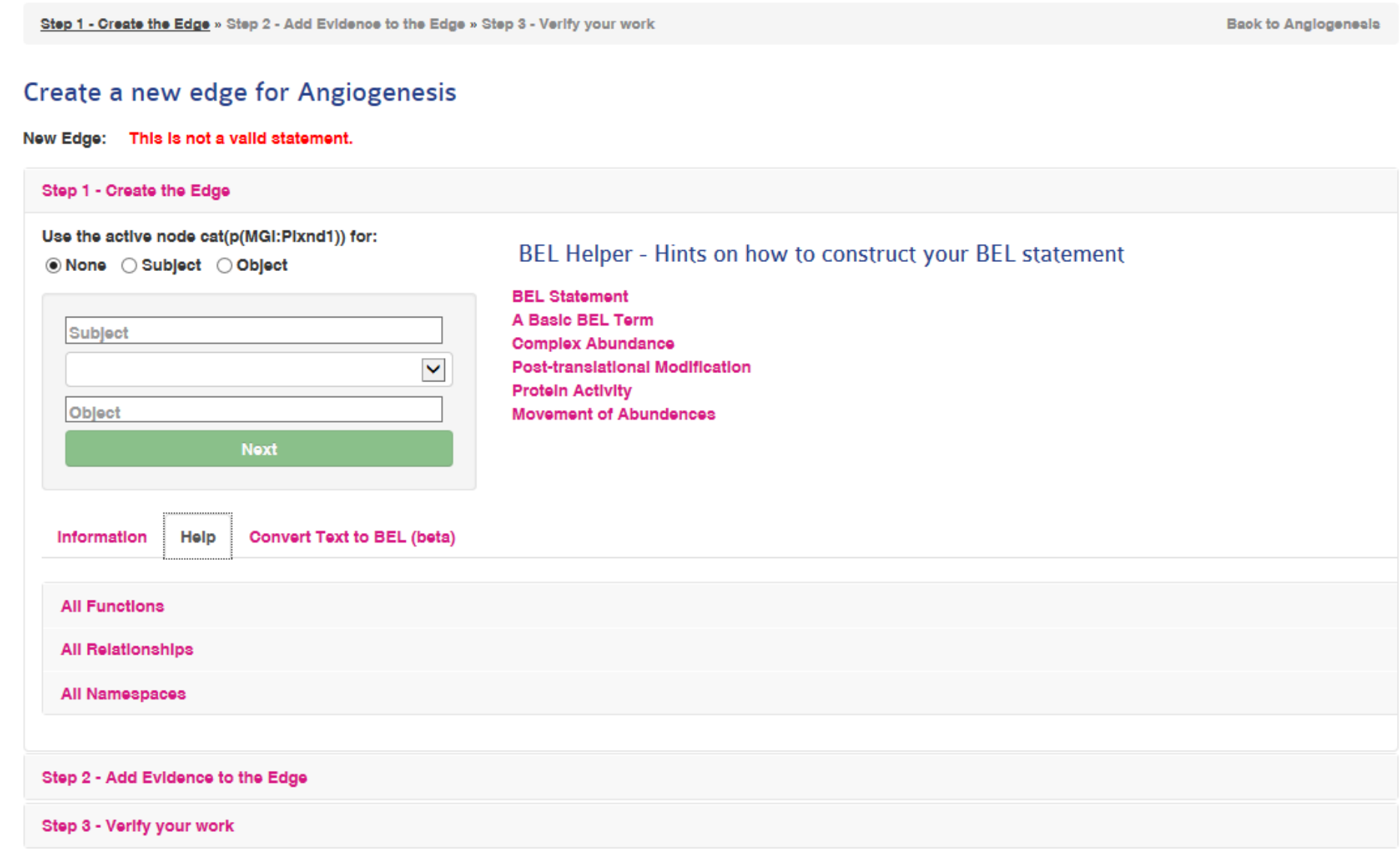
Network Verification - Voting



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

https://sbvimprover.com/challenge-3/challenge

Network Verification – Extension



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Network Verification – Outcome

Activity during the open phase (04/2014 – 05/2015)

9,286 votes

2,225 new pieces of evidence

1,289 new edges

1,000 new nodes

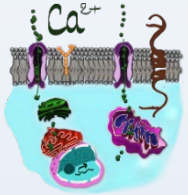
Participants

173 participants from 26 countries

20 Best Performers invited to the Jamboree in Barcelona, Spain with a travel bursary.

Jamboree Session on…

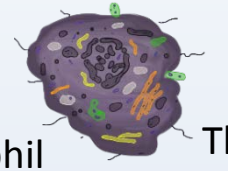
Calcium signaling



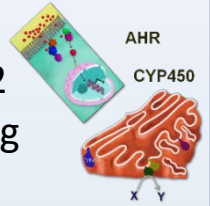
Epigenetics

Oxidative Stress

Necroptosis

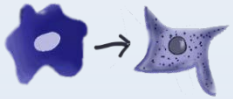


Xenobiotic Metabolism



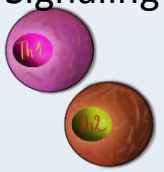
Macrophage signaling

Senescence



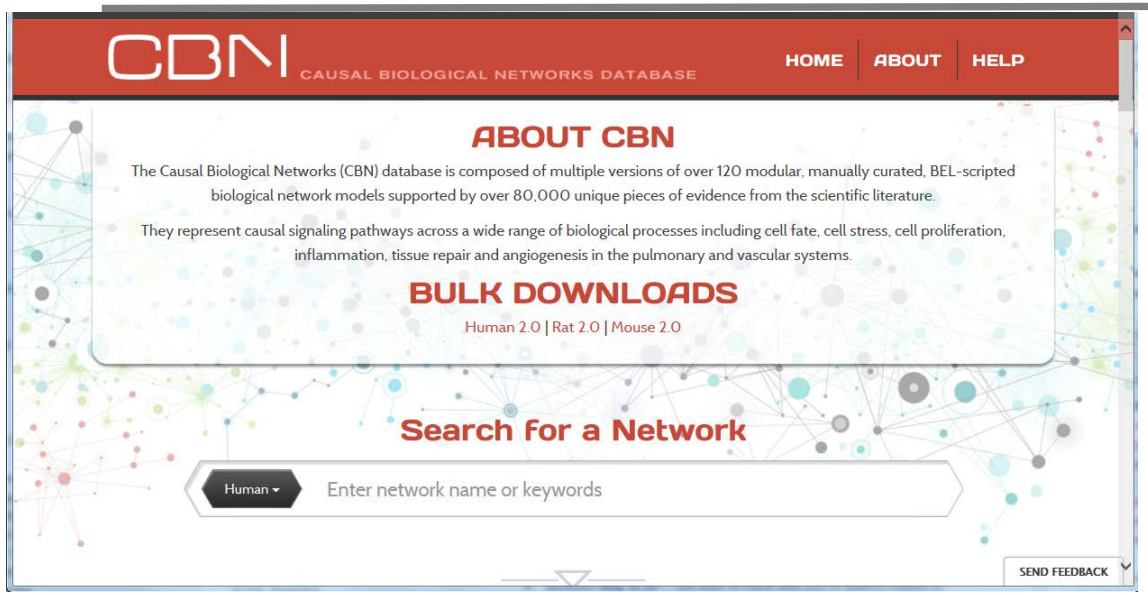
Neutrophil Signaling

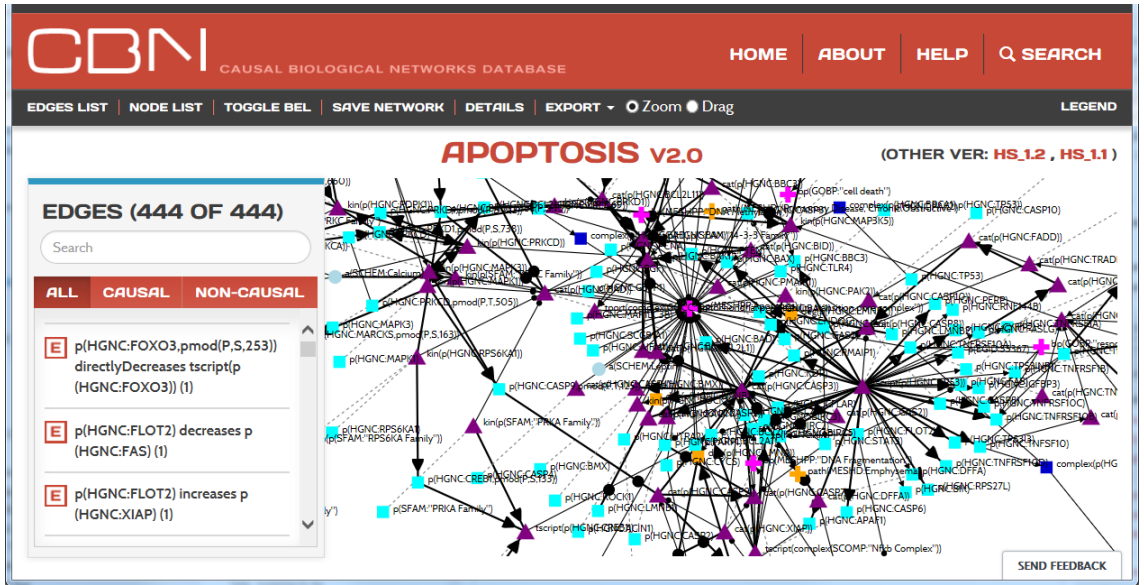
Th1-Th2 Signaling

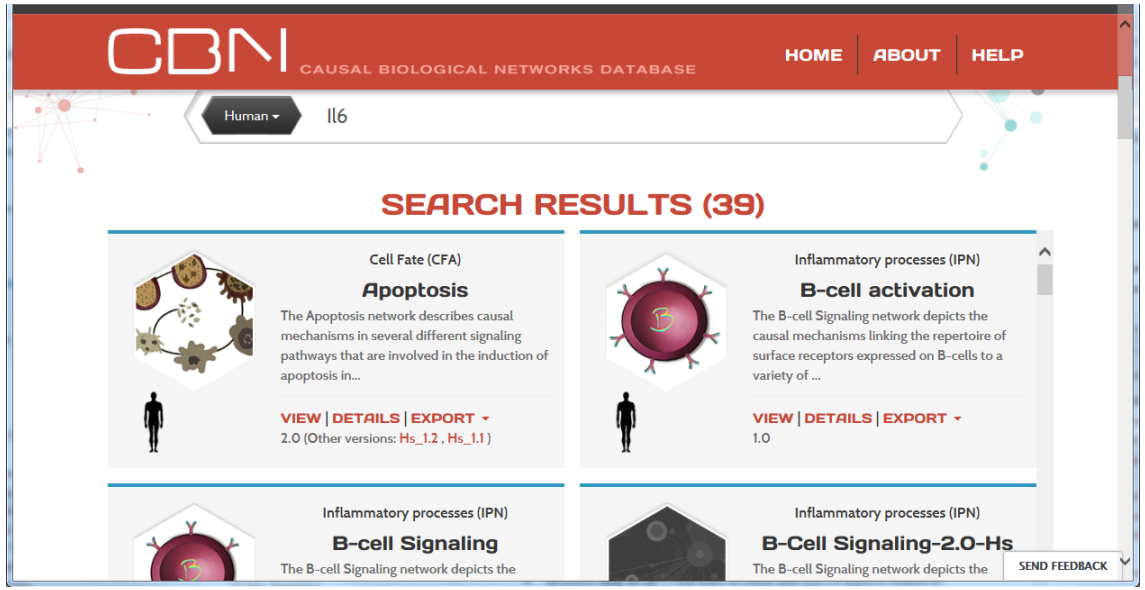


PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Network Verification – Network Sharing







Boue, S., Talikka, M., Westra, JW., et al. (2015), Database (Oxford). 2015 Apr 17;2015, Causal biological network database: a comprehensive platform of causal biological network models focused on the pulmonary and vascular systems.

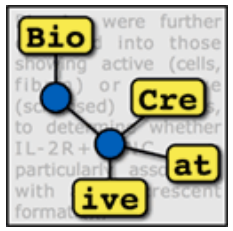
PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

Summary

- Our approach in using systems toxicological assessment requires computational and knowledge-driven analysis of data from various experimental techniques
- Knowledge is extracted from scientific articles and converted into a human and computer-readable format: BEL
- **BELIEF is supporting the automated extraction of knowledge as well as the manual curation and outputs in BEL**
- BEL networks can be verified by the crowd using the sbvIMPROVER Network Verification Challenge (NVC)
- Reviewed and verified networks are shared in the Causal Biological Networks Database (CBN)

Madan, S., Hodapp, S., Senger, P., et al. (2016) Database (Oxford)., 2016, The BEL Information Extraction Workflow (BELIEF): Evaluation in the BioCreative V BEL and IAT track (accepted)

Acknowledgements



PMI high performance computing