



PMI SCIENCE
PHILIP MORRIS INTERNATIONAL

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

Sam Ansari

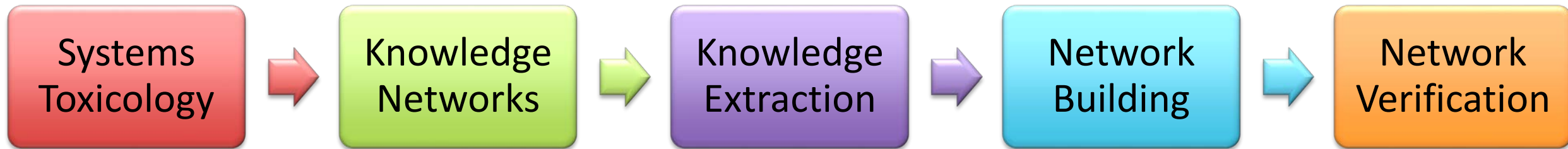
*Philip Morris International R&D, Philip Morris Products S.A.
(part of Philip Morris International group of companies)*

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

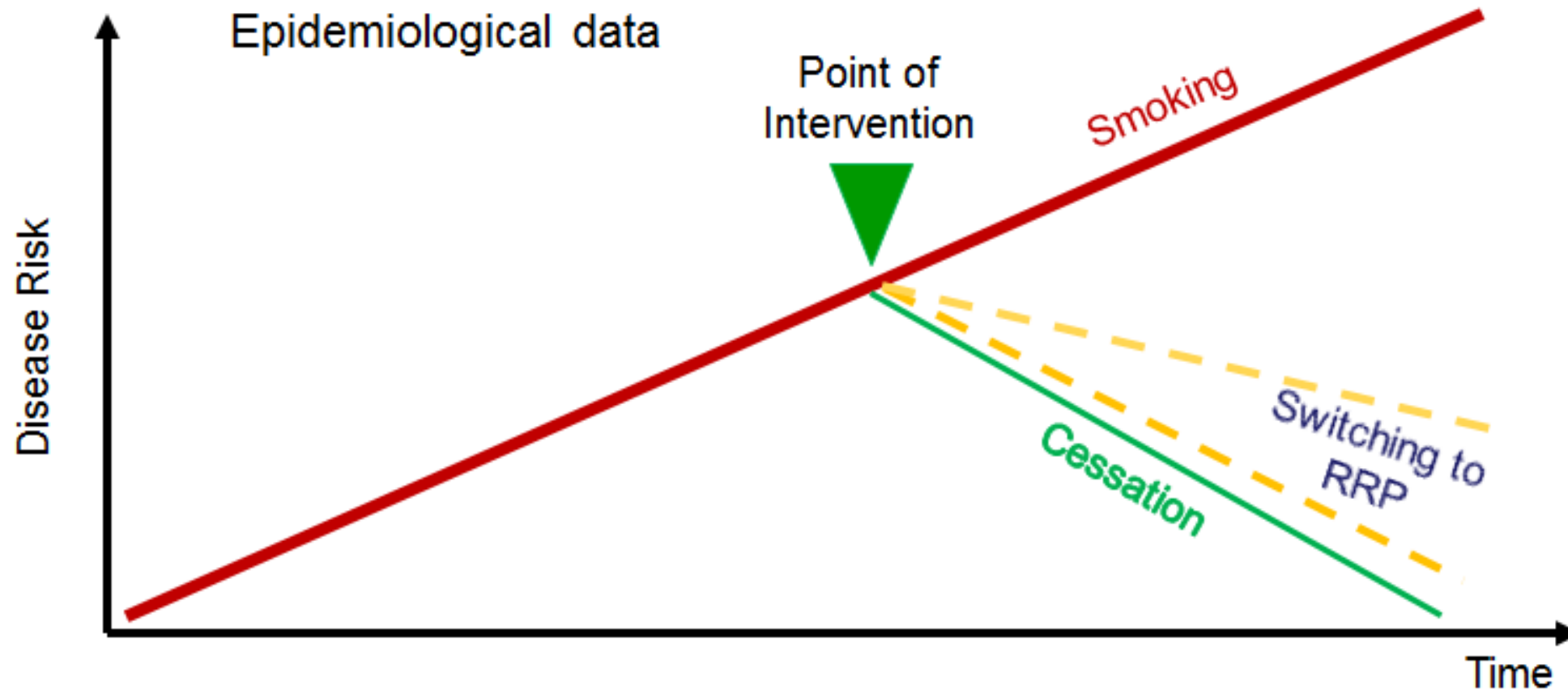
Smoking causes serious diseases such as cardiovascular diseases, lung cancer, and chronic obstructive pulmonary disease.

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



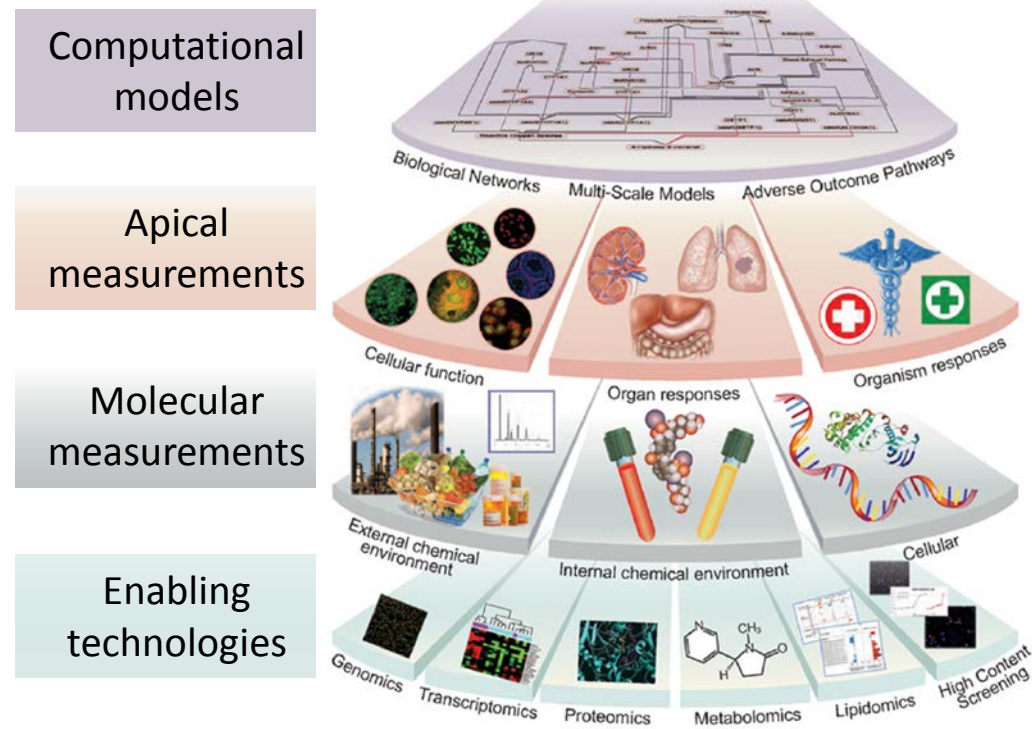
Note: Reduced-Risk Products ("RRPs") is the term we use to refer to products that have the potential to reduce individual risk and population harm in comparison to smoking combustible cigarettes

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 throughput methods with advanced computational methods
- enables the shift to a new paradigm for risk assessment which is the future toxicology of the 21st century (Product Assessment)





Systems Toxicology – Assessment Approach

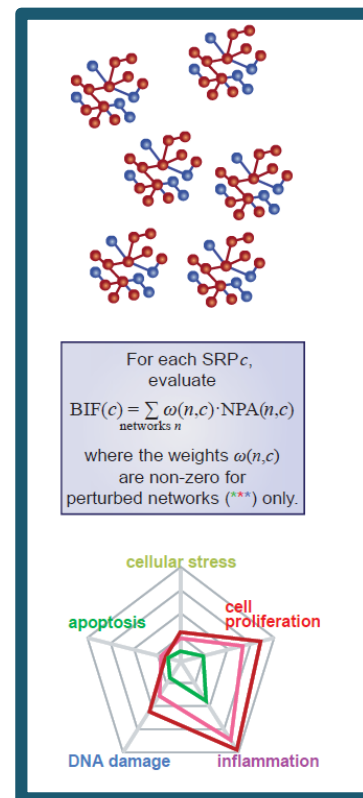
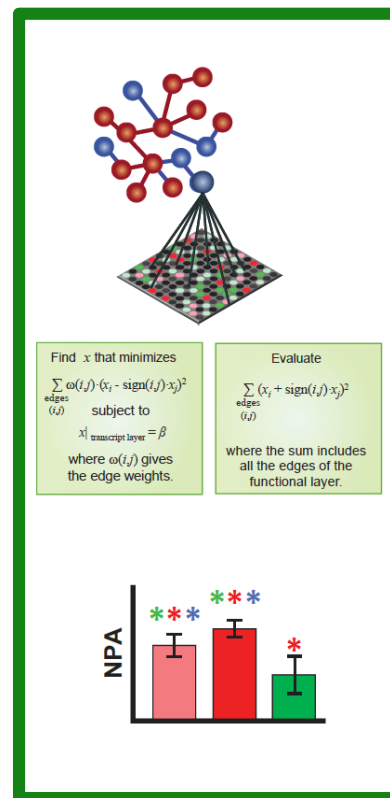
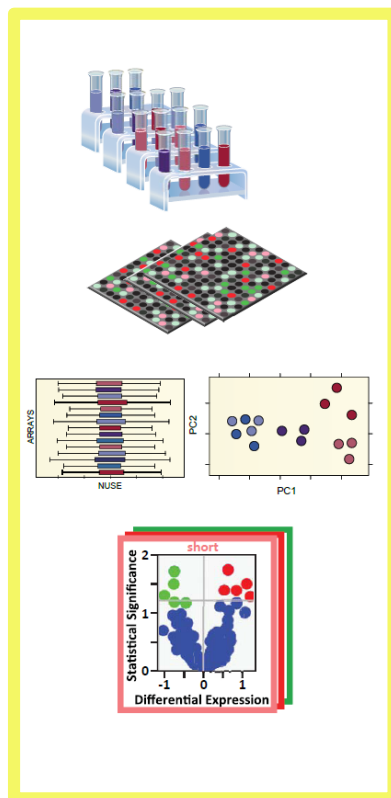
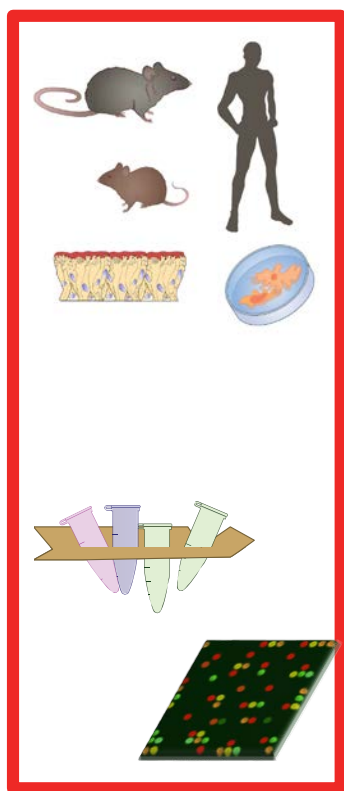
Experimental
data production

Compute
systems
response profiles

Identify
perturbed
biological
networks

Compute
network
perturbation
amplitudes

Compute
product
biological impact



Knowledge Networks

From Literature to Computable Knowledge

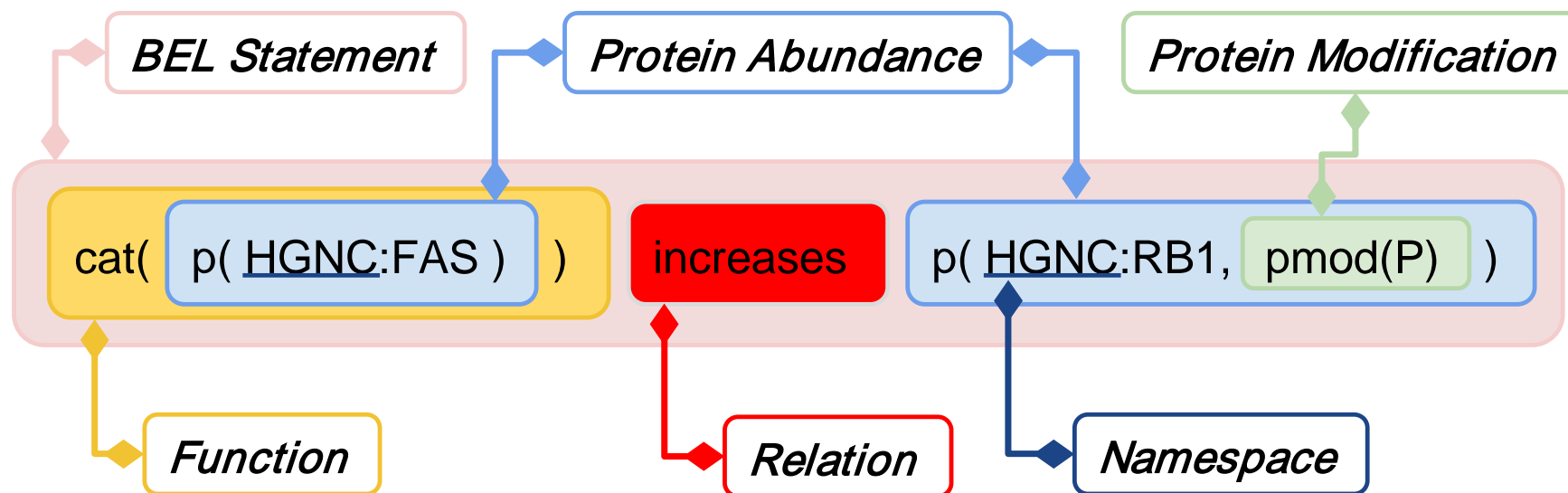




Knowledge Networks - Biological Expression Language (BEL)



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

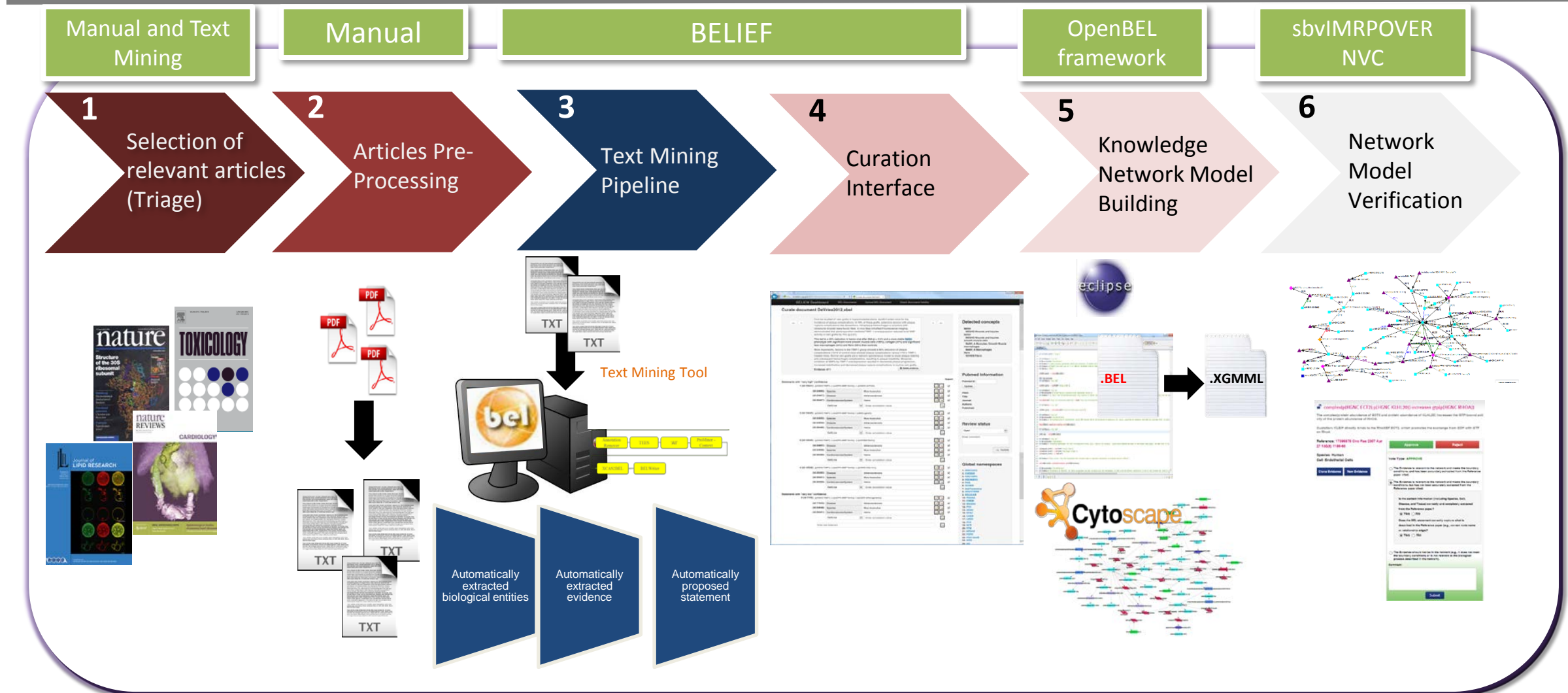
SET Ti ssue = “jurkat cells”

Knowledge Networks

From Literature to Computable Knowledge

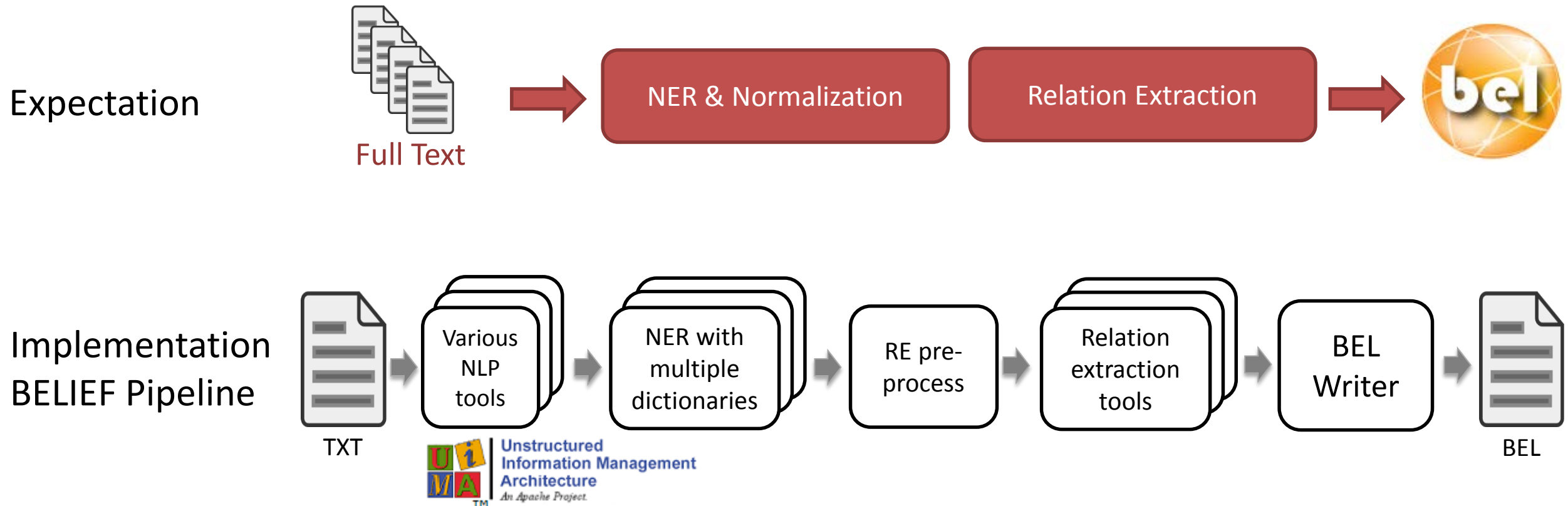


Knowledge Extraction - Overview





Knowledge Extraction - BELIEF Pipeline





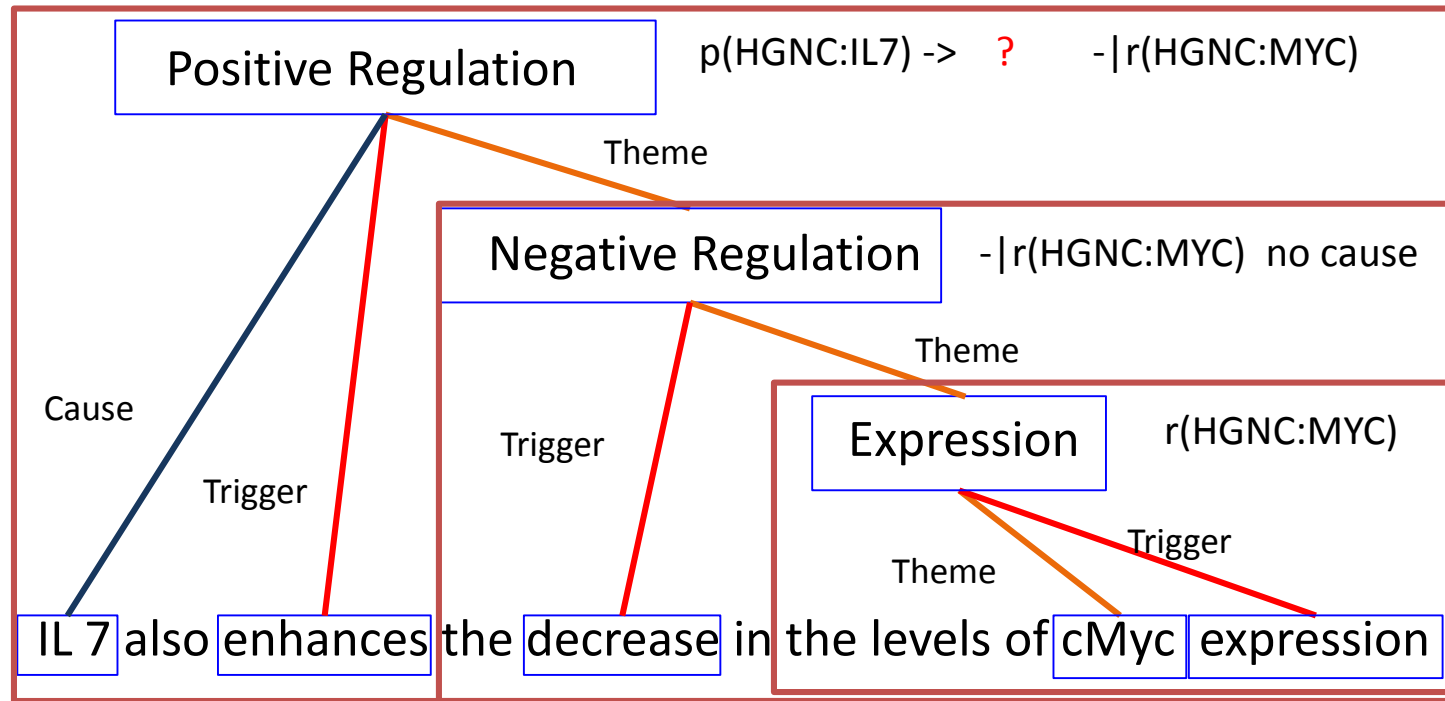
Knowledge Extraction - Named Entity Recognition

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

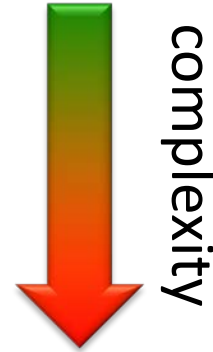
IL7 - cMyc Relation: Yes

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

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

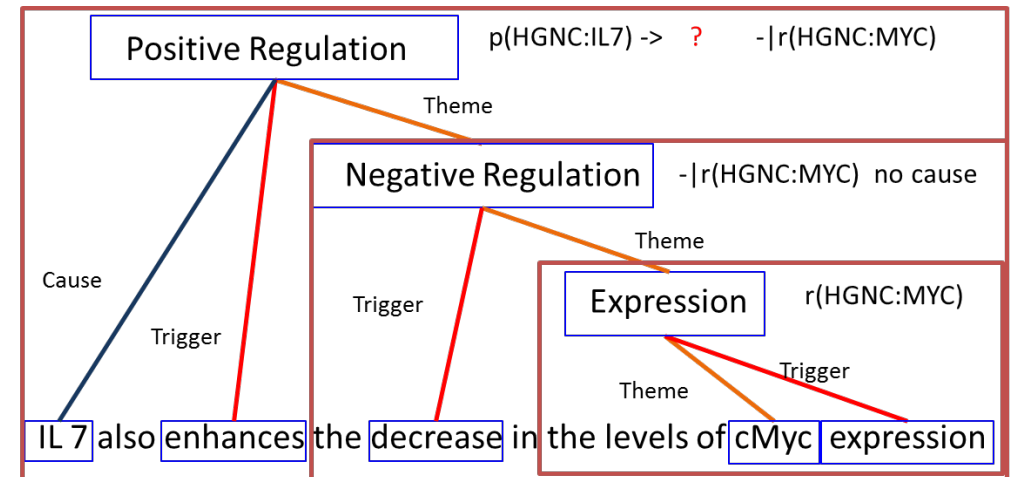
BELIEF Pipeline - Limitations

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



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)



Knowledge Extraction - Curation Interface

Curate doc
sirtuin 1.

[Return to document list](#)

Detected concepts:

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

SIRT1 knockdown caused a slight yet highly significant decrease in LXAlpha expression in both fed and fasted **livers**.

However, SIRT1 knowndown decreased expression of PGC-1beta only the fasted state.

Evidence: 1/2

Delete evidence

Create table evidence

BEL statement

Context annotation

Statements with "very low" confidence:

1 (id:44394): p(FIXME) -| (p(HGNC:SIRT1) -| r(MGI:Nr1h3))

(id:90567):

MeSHAnatomy

Liver

CellLine

Enter annotation value

Enter new statement

Add/edit/delete/export

BEL namespace search

Document PubMed
metadata

Concepts

SIRT1

MGI:Sirt1

HGNC:SIRT1

RGD:Sirt1

LXAlpha

MGI:Nr1h3

RGD:Nr1h3

livers

MeSHAnatomy:Liver

Browse namespaces

type e.g. HGNC:CDK1

Pubmed Information

PubmedId: 17646659

Update

PMID: 17646659

Abstract: Fasting-dependent glucose and lipid
metabolic response through hepatic sirtuin 1.

Journal: Proceedings of the National
Academy of Sciences of the United States of
America, Vol. 104; Iss. 31

Authors: Joseph T Rodgers, Pere Puigserver

Published: 2007-07-31 00:00:00 CEST

BELIEW Dashboard

[Home](#)
[BEL-Documents](#)
[Upload BEL-Document](#)
[Projects ▾](#)
[Documents ▾](#)

<

List
Create

Search:

Show

10

entries

Search:

ID	Statement	Evidence ID	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.	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
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.	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
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.	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
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.	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
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.	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
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.	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Showing 1 to 2 of 2 entries

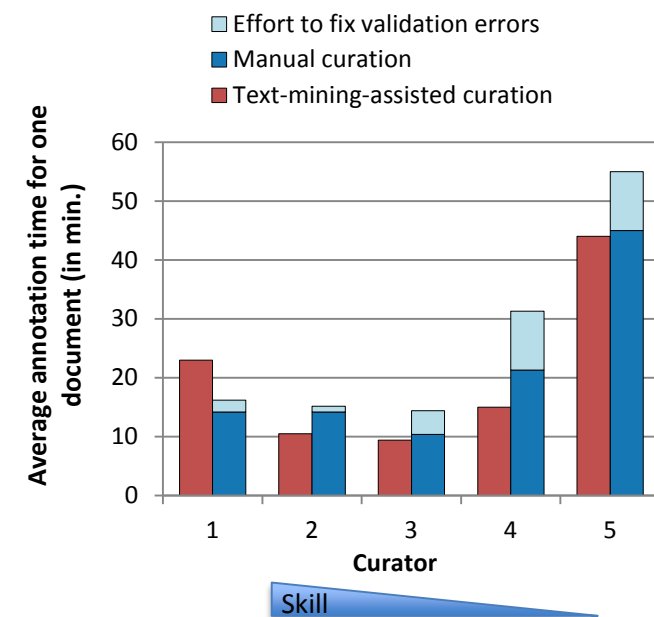
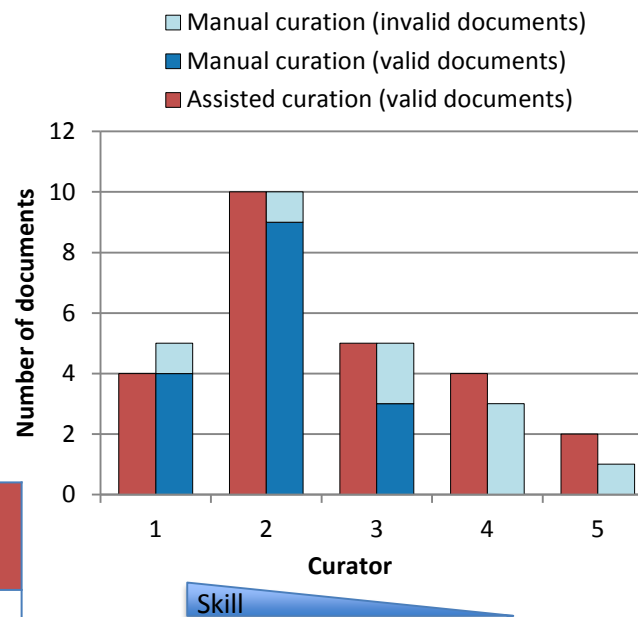
Previous

1

Next

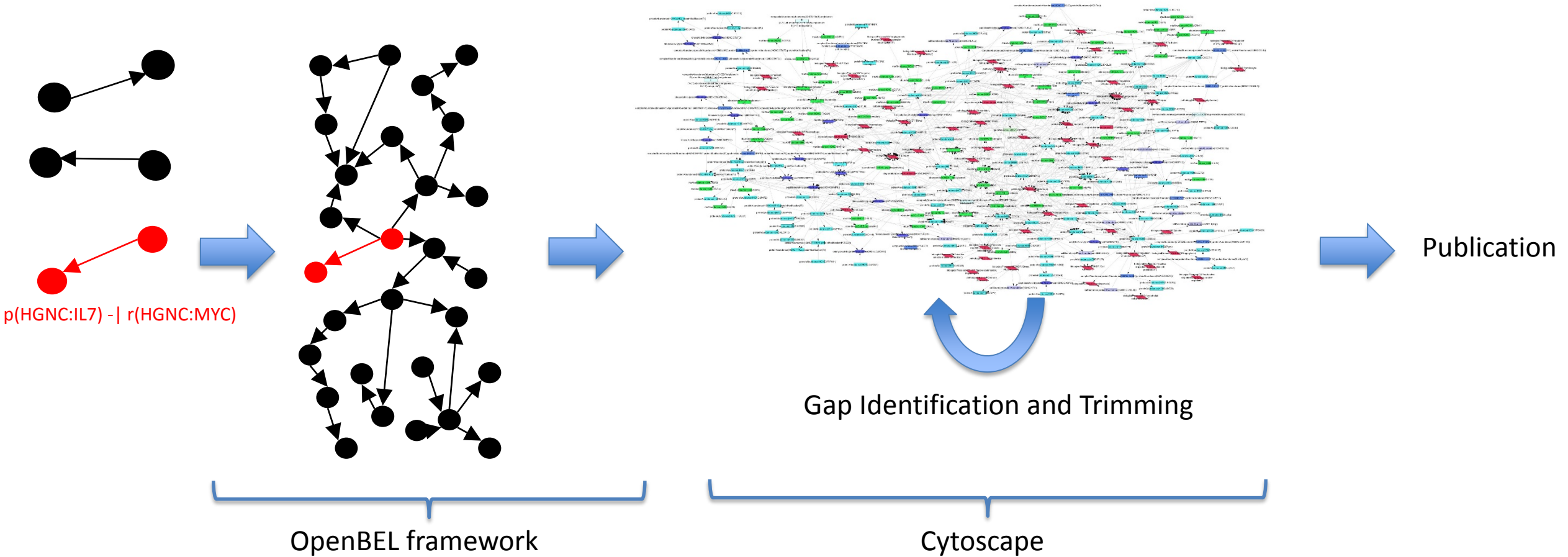
BELIEF Dashboard - Performance and Impression

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



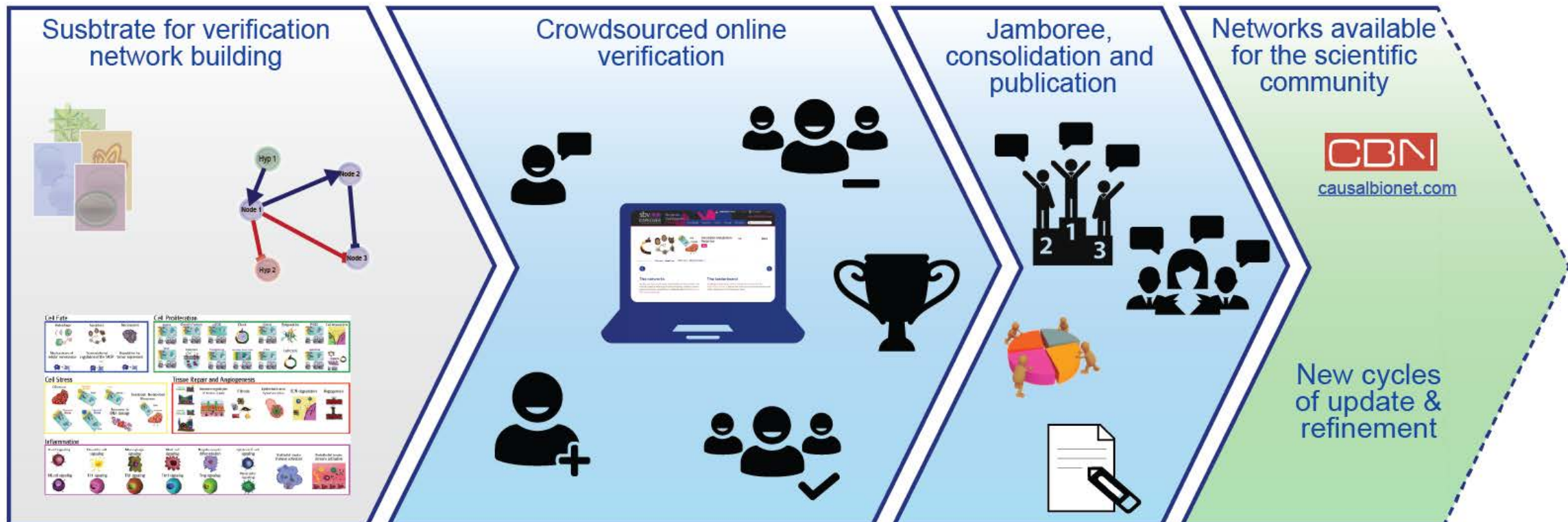


Network Building





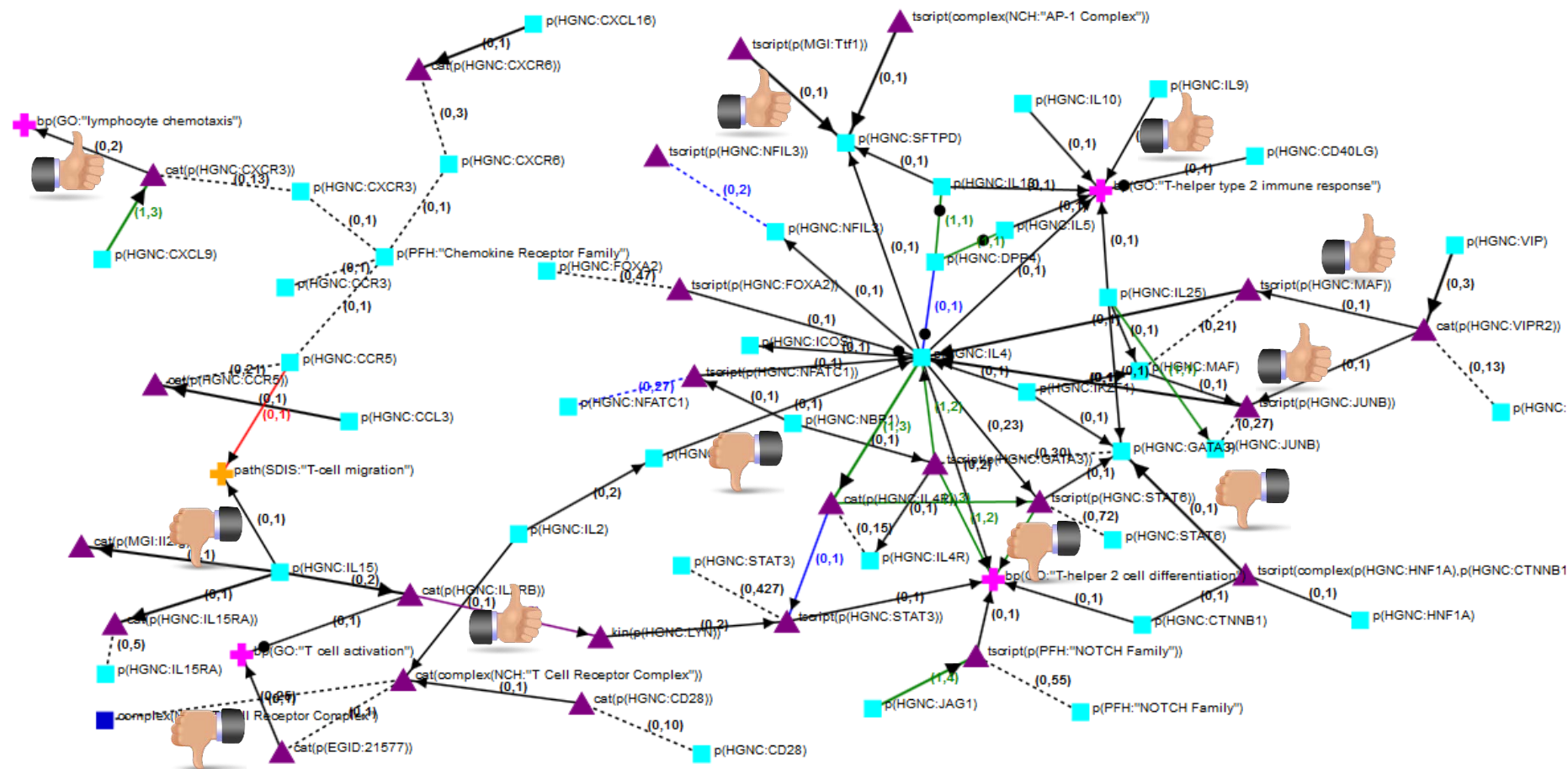
Crowd Verification - Overview





Network Verification - Concept

Vote on evidence, create new edges, add missing nodes



Network Verification - Selection of Sub-network





Network Verification - Review of Triples

Th2 Signaling

Download Network Locked Reset Display Network Details Network Activity Legend Comments Nodes Edges

Edge Comment

A Normal text

Bold *Italic* Underline

Enter text ...

Save

Evidence Detail

tscript(p(HGNC:GATA3)) increases p(HGNC:IL4)

The transcriptional activity of the protein abundance of GATA3 increases the protein abundance of IL4.

Quotation: The GATA3 transcription factor induces IL-4 and IL-2 production in Th1, Th2 and Th17 cells.

Reference: 18792410 Eur J Immunol 2008 Sep 15 38(9): 2573-2586

Species: Mouse
Cell: t-cell

Create Evidence New Evidence

This Network is locked and you cannot Approve or Reject this Evidence at this time

Selected Edge

tscript(p(HGNC:GATA3)) increases p(HGNC:IL4)

Comment New Evidence

Evidence (2)

✓ 4.0 Mouse t
✓ 4.0 Mouse t host t

Votes

Created By	Voted	Reason	Comment
svgur	Accepted	The Evidence is relevant to the network and meets the boundary conditions, and has been accurately extracted from the Reference paper cited.	
sashapro2005	Accepted	The Evidence is relevant to the network and meets the boundary conditions, and has been accurately extracted from the Reference paper cited.	
Diego.Morales	Accepted	The Evidence is relevant to the network and meets the boundary conditions, and has been accurately extracted from the Reference paper cited.	
Ilonabinenbaum	Accepted	The Evidence is relevant to the network and meets the boundary conditions, and has been accurately extracted from the Reference paper cited.	

Edges (15)

All Causal Non-Causal

Search

- tscript(p(HGNC:GATA3)) directly increases p(HGNC:IL4)
- p(HGNC:GATA3) increases p(HGNC:IL4)
- tscript(p(HGNC:GATA3)) increases p(HGNC:IL4)
- p(HGNC:GATA3) decreases p(HGNC:IL4)
- tscript(p(HGNC:GATA3)) decreases p(HGNC:IL4)
- p(HGNC:IL4) increases bp(GO:TT-helper 2 cell differentiation)
- p(HGNC:IL4) increases bp(GO:TT-helper type 2 immune response)
- p(HGNC:IL4) increases tscript(p(HGNC:STAT6))
- p(HGNC:IL4) directly increases card(p(HGNC:IL4R))
- p(HGNC:IL4) increases p(HGNC:NFKB)
- p(HGNC:IL4) increases p(HGNC:STAT6)
- p(HGNC:IL4) increases p(HGNC:GATA3)
- tscript(p(HGNC:GATA3)) increases p(HGNC:IL4)
- tscript(p(HGNC:GATA3)) directly increases p(HGNC:IL4)
- p(HGNC:IL4) increases p(HGNC:IL4)



Network Verification - Voting

 complex(p(HGNC:ECT2),p(HGNC:KLHL20)) increases gtp(p(HGNC:RHOA))

The complex(protein abundance of ECT2 and protein abundance of KLHL20) increases the GTP-bound activity of the protein abundance of RHOA.

Quotation: KLEIP directly binds to the RhoGEF ECT2, which promotes the exchange from GDP with GTP on RhoA

Reference: 17395875 Circ Res 2007 Apr
27 100(8) 1155-63

Species Human
Cell Endothelial Cells

Clone Evidence

New Evidence

Approve

Reject

Vote Type: **APPROVE**

- ☐ The Evidence is relevant to the network and meets the boundary conditions, and has been accurately extracted from the Reference paper cited.
- ☒ The Evidence is relevant to the network and meets the boundary conditions, but has not been accurately extracted from the Reference paper cited.

Is the context information (including Species, Cell, Disease, and Tissue) correctly and completely extracted from the Reference paper?

☒ Yes ☐ No

Does the BEL statement correctly capture what is described in the Reference paper (e.g., correct node name or relationship edge)?

☒ Yes ☐ No

- ☐ The Evidence should not be in the network (e.g., it does not meet the boundary conditions or is not relevant to the biological process described in the network).

Comment

Submit



Network Verification - Extension

[Step 1 - Create the Edge](#) » [Step 2 - Add Evidence to the Edge](#) » [Step 3 - Verify your work](#)

[Back to Angiogenesis](#)

Create a new edge for Angiogenesis

New Edge: **This is not a valid statement.**

Step 1 - Create the Edge

Use the active node `cat(p(MGI:Plxnd1))` for:

☒ None ☐ Subject ☐ Object

Subject



Object

Next

BEL Helper - Hints on how to construct your BEL statement

BEL Statement

A Basic BEL Term

Complex Abundance

Post-translational Modification

Protein Activity

Movement of Abundances

[Information](#)

[Help](#)

[Convert Text to BEL \(beta\)](#)

[All Functions](#)

[All Relationships](#)

[All Namespaces](#)

[Step 2 - Add Evidence to the Edge](#)

[Step 3 - Verify your work](#)



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

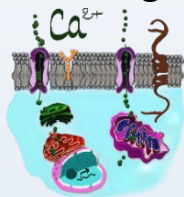
Participants

173 participants
from 26 countries

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

Jamboree Session on...

Calcium signaling



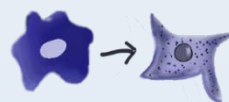
Macrophage
signaling



Epigenetics



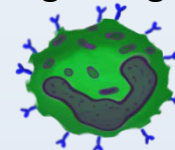
Senescence



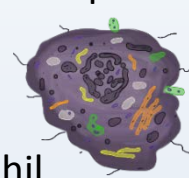
Oxidative
Stress



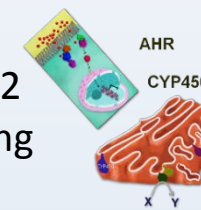
Neutrophil
Signaling



Necroptosis



Xenobiotic
Metabolism



Th1-Th2
Signaling





Network Verification - Network Sharing

CBNI CAUSAL BIOLOGICAL NETWORKS DATABASE

HOME ABOUT HELP

ABOUT CBN

The Causal Biological Networks (CBN) database is composed of multiple versions of over 120 modular, manually curated, BEL-scripted biological network models supported by over 80,000 unique pieces of evidence from the scientific literature.

They represent causal signaling pathways across a wide range of biological processes including cell fate, cell stress, cell proliferation, inflammation, tissue repair and angiogenesis in the pulmonary and vascular systems.

BULK DOWNLOADS

Human 2.0 | Rat 2.0 | Mouse 2.0

Search for a Network

Human Enter network name or keywords

SEND FEEDBACK

CBNI CAUSAL BIOLOGICAL NETWORKS DATABASE

HOME ABOUT HELP Q SEARCH

EDGES LIST NODE LIST TOGGLE BEL SAVE NETWORK DETAILS EXPORT Zoom Drag LEGEND

APOPTOSIS v2.0 (OTHER VER: HS_1.2, HS_1.1)

EDGES (444 OF 444)

Search

ALL CAUSAL NON-CAUSAL

- E** p(HGNC:FOXO3,pmod(P,S,253)) directlyDecreases tscript(p(HGNC:FOXO3)) (1)
- E** p(HGNC:FLOT2) decreases p(HGNC:FAS) (1)
- E** p(HGNC:FLOT2) increases p(HGNC:XIAP) (1)

SEND FEEDBACK

CBNI CAUSAL BIOLOGICAL NETWORKS DATABASE

HOME ABOUT HELP

Human 116

SEARCH RESULTS (39)

Cell Fate (CFA)

Apoptosis

The Apoptosis network describes causal mechanisms in several different signaling pathways that are involved in the induction of apoptosis in...

VIEW | DETAILS | EXPORT 2.0 (Other versions: Hs_1.2, Hs_1.1)

Inflammatory processes (IPN)

B-cell activation

The B-cell Signaling network depicts the causal mechanisms linking the repertoire of surface receptors expressed on B-cells to a variety of ...

VIEW | DETAILS | EXPORT 1.0

Inflammatory processes (IPN)

B-cell Signaling

The B-cell Signaling network depicts the

VIEW | DETAILS | EXPORT

Inflammatory processes (IPN)

B-Cell Signaling-2.0-Hs

The B-cell Signaling network depicts the

VIEW | DETAILS | EXPORT

SEND FEEDBACK

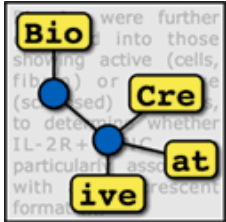
Boue, S., Talikka, M., Westra, JW., et al. (2015), Database (Oxford). 2015 Apr 17;2015, Causal biological network database: a comprehensive platform of causal biological network models focused on the pulmonary and vascular systems.

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)



Acknowledgements



PMI high performance
computing

