Refined JST Thesaurus Extended with Data from other Open Life Science Data Sources

Tatsuya Kushida[1], Yuka Tateisi[1], Takeshi Masuda[2], Katsutaro Watanabe[3], Katsuji Matsumura[3], Takahiro Kawamura[3], Kouji Kozaki[2], and Toshihisa Takagi[1,4]

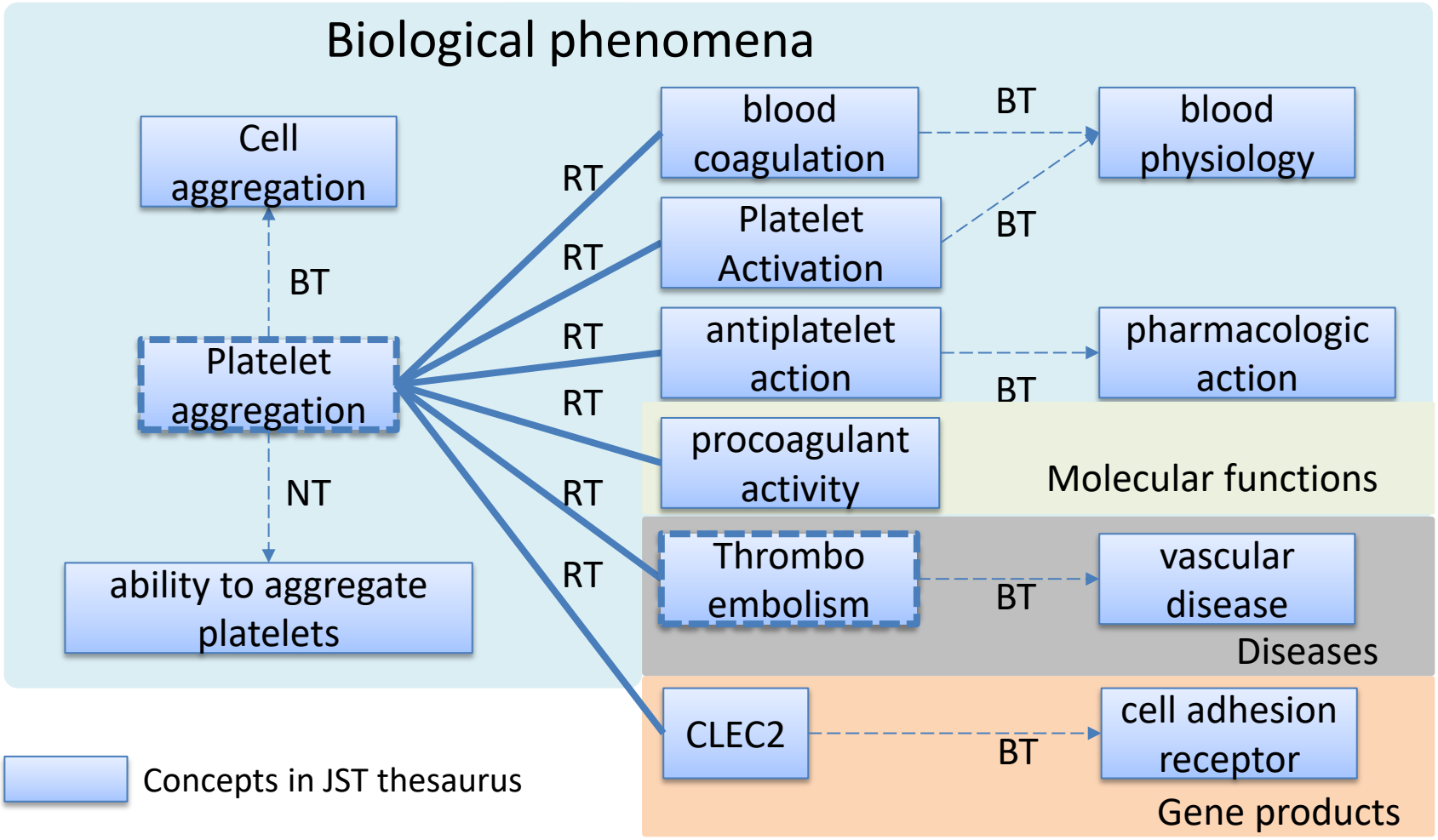
1 National Bioscience Database Center, Japan Science and Technology Agency, Tokyo, Japan
2 The Institute of Scientific and Industrial Research, Osaka Univ., Ibaraki, Japan
3 Dept. of Information Planning, Japan Science and Technology Agency, Tokyo, Japan
4 Dept. Biological Sciences, Grad. School of Science, The Univ. of Tokyo, Tokyo, Japan

JIST2017, 10 - 12 November 2017
Gold Coast, Australia

Japan Science and Technology (JST) thesaurus

- It is developed and provided by Japan Science and Technology Agency.
- It is mainly used for indexing scientific literatures.
- All of the terms are written in both of English and Japanese languages.
- It contains approximately 245,000 concepts including **90,000 life science concepts**.
  - Gene products (e.g. CLEC2)
  - Drugs (e.g. Gefitinib)
  - Biological phenomena (e.g. platelet aggregation)
  - Diseases (e.g. Thromboembolism)
  - Anatomy (e.g. cartilage)
  - ...

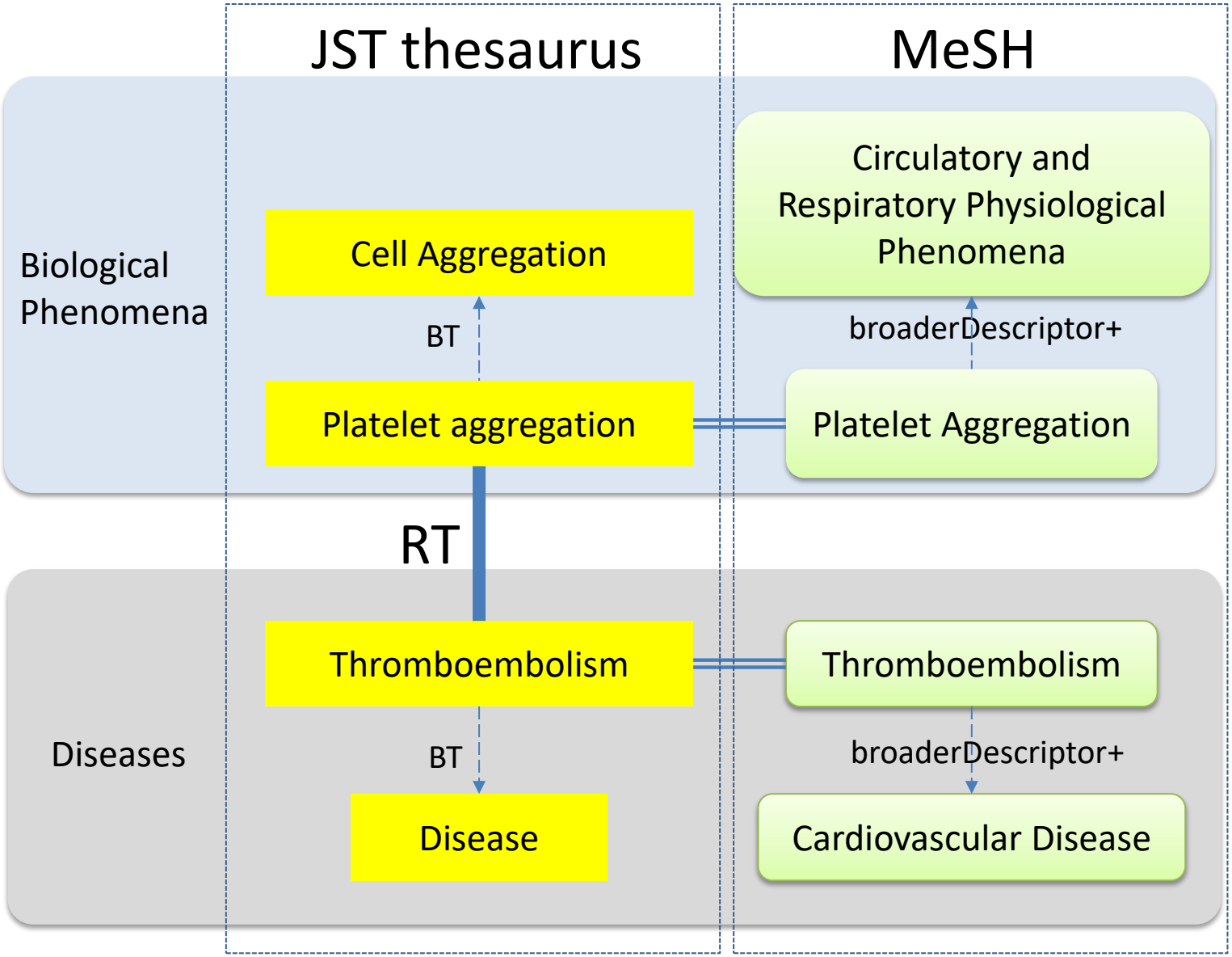
Structure of JST thesaurus



BT: Broader term NT: Narrower term RT: Related term

Comparison between JST thesaurus and MeSH

The advantages of JST thesaurus



+: A path of length one or more

Problems of the original JST thesaurus (1/2)

Problem #1

The JST thesaurus has only three kinds of simple relations (BT, NT, and RT).

Therefore, we cannot describe the following relationships,

- Diseases, and the preceding biological phenomena
- Disease states, and the succeeding ones
- Diseases, and gene products regulating them
- etc.

Solution to Problem #1

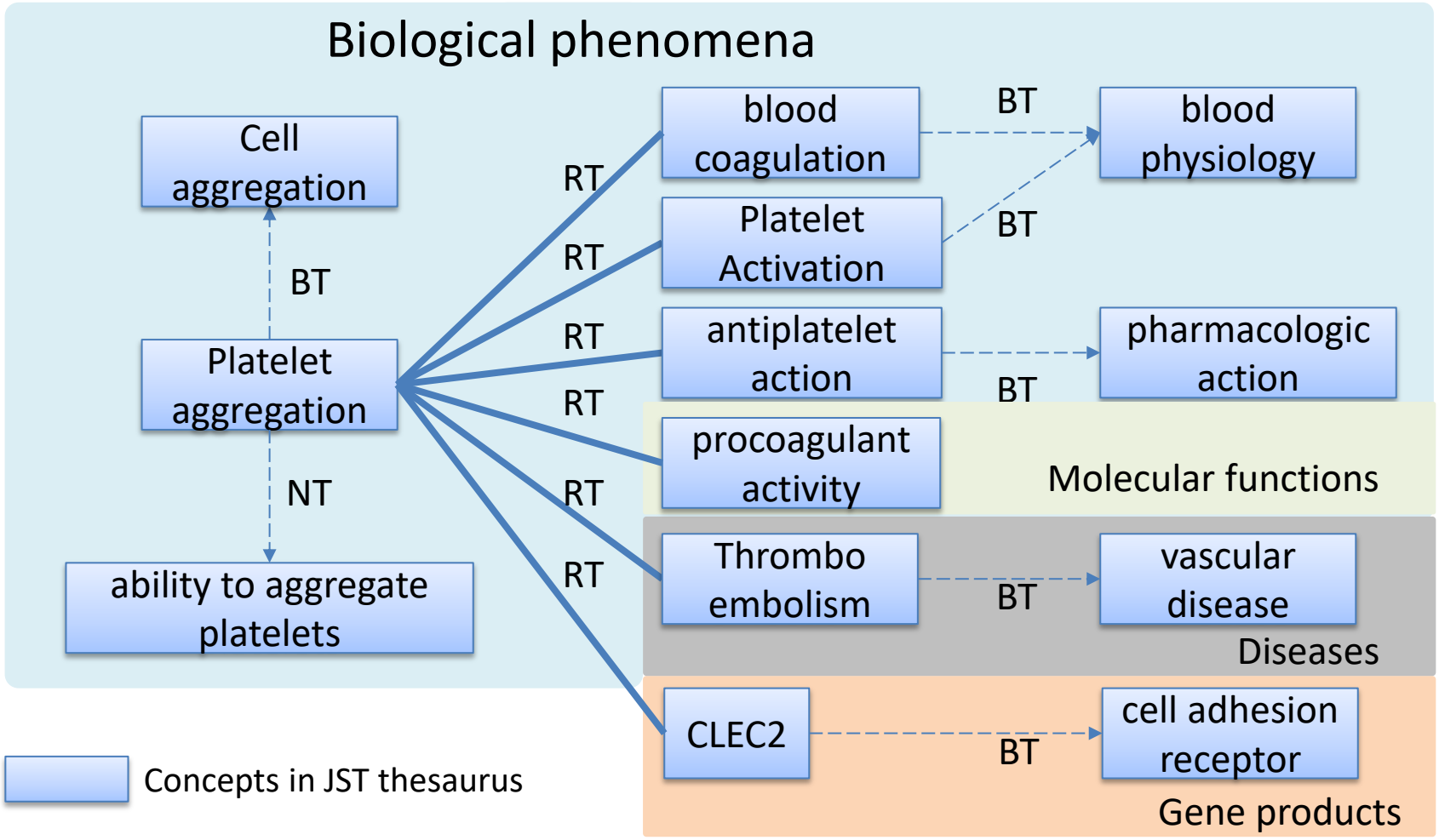
We have **sub-classified RTs** by using thirty-one kinds of relations such as “has function”, and we name the thesaurus “**Refined JST thesaurus**”.

Thirty-one relations used for the RT sub-classifying

rdfs:subClassOf
skos:narrower (has narrower)
sio:SIO_000028 (has part)
sio:SIO_000068 (is part of)
sio:SIO_000218 (is quality of)
sio:SIO_000217 (has quality)
sio:SIO_000226 (is function of)
sio:SIO_000225 (has function)
sio:SIO_000123 (antonym)
sio:SIO_000228 (has role)
sio:SIO_000227 (is role of)
sio:SIO_001279 (has phenotype)
nbdc:isPhenotypeOf
sio:SIO_001154 (regulates)
sio:SIO_001155 (is regulated by)
xkos:succeeds
xkos:precedes
sio:SIO_000657(is transformed from)
sio:SIO_000655 (transforms into)
sio:SIO_000203 (is connected to)
sio:SIO_000365 (is creator of)
sio:SIO_000364 (has creator)
sio:SIO_000145 (is location of)
sio:SIO_000061 (is located in)
obo:RO_0002234 (has output)
obo:RO_0002353 (output of)
sio:SIO_000064 (is provider of)
sio:SIO_000066 (has provider)
sio:SIO_000122 (synonym)
sio:SIO_000283 (is similar to)
skos:related (RT)

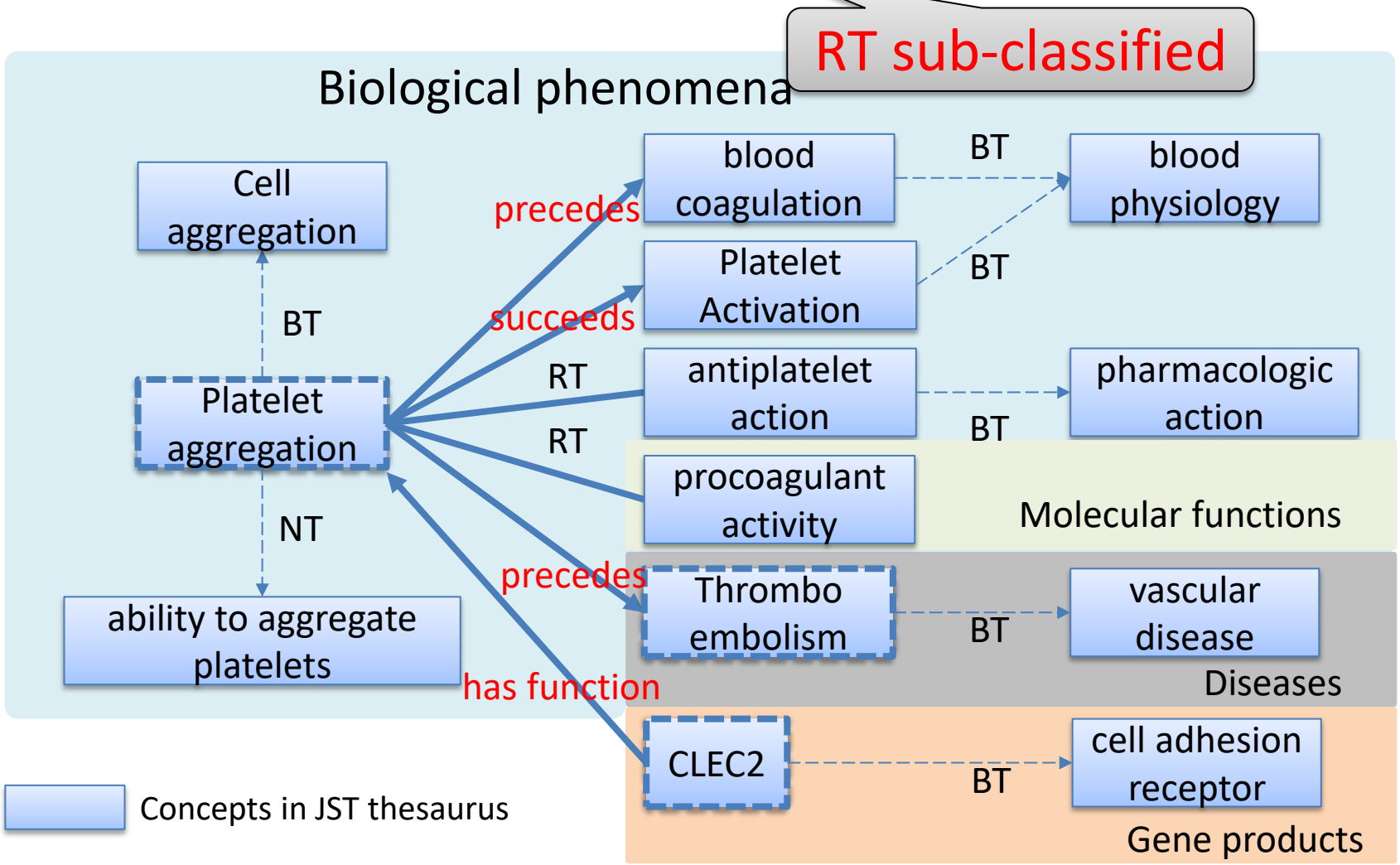
Semanticscience Integrated Ontology (**SIO**)
sio: http://semanticscience.org/resource/

Structure of original JST thesaurus



BT: Broader term NT: Narrower term RT: Related term

Structure of refined JST thesaurus



BT: Broader term NT: Narrower term RT: Related term

Problems of the original JST thesaurus (2/2)

Problem #2

The JST thesaurus is designed to collect concepts broadly and shallowly.

For example, the JST thesaurus includes **two** “fibrinolysis” related gene products. On the other hand, Gene Ontology includes the **twenty-seven** related gene products .

Solution to Problem #2

We extend the refined JST thesaurus by incorporating information from other ontologies, thesauri, databases and LOD.

Aim of this study

Aim

- To update the refined JST thesaurus using thirty-five relations including **new relations**
- To evaluate the effectiveness of using sub-classified relations (refined JST thesaurus) by comparing **networks** (graphs) constructed from the refined JST thesaurus with those from the original JST thesaurus
- To integrate the refined JST thesaurus with other **ontologies, thesauri, and LOD**.

Updating the refined JST thesaurus using thirty-five relations including new relations

List of newly added relations

Relations	Definitions
sio:SIO_000230 (has input)	Relation between a **material** and a ***process*** related to it as the input to the **molecular mechanism**
sio:SIO_000231 (is input in)	Reverse relation of "has input"
sio:SIO_000132 (has participant)	Relation between a **material** or **process** and a ***process*** related to it in the **molecular mechanism**
sio:SIO_000062 (is participant in)	Reverse relation of "has participant"

Red bold underlined letter: Domain of the relation
Blue Italic underlined letter: Range of the relation

Evaluating the effectiveness of using sub-classified relations by comparing networks (graphs)

&

Example: Classifying diseases using sub-classified relations (refined JST thesaurus)

Storing the RDF data in the triple store, and constructing knowledge graphs

Refined JST thesaurus

Convert to RDF (Turtle)

I

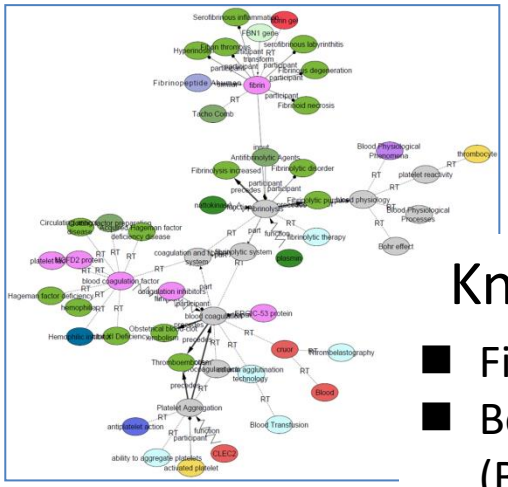
Ho - zo 法造 an Environment for Building/Using Ontologies www.hozo.jp

```
## Example ##
jstid:20100607047002124 7
 sio:SIO_000225 jstid:20090603419802747 0 .
```

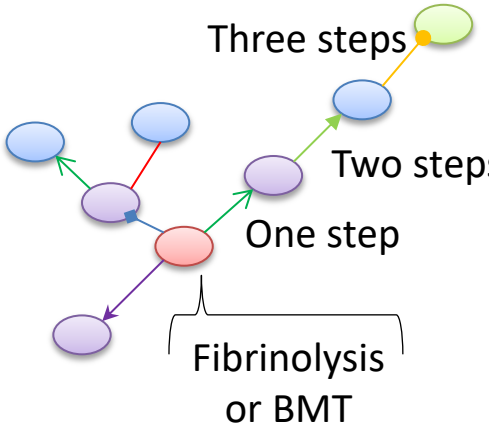
II

Triple store (Virtuoso)

III



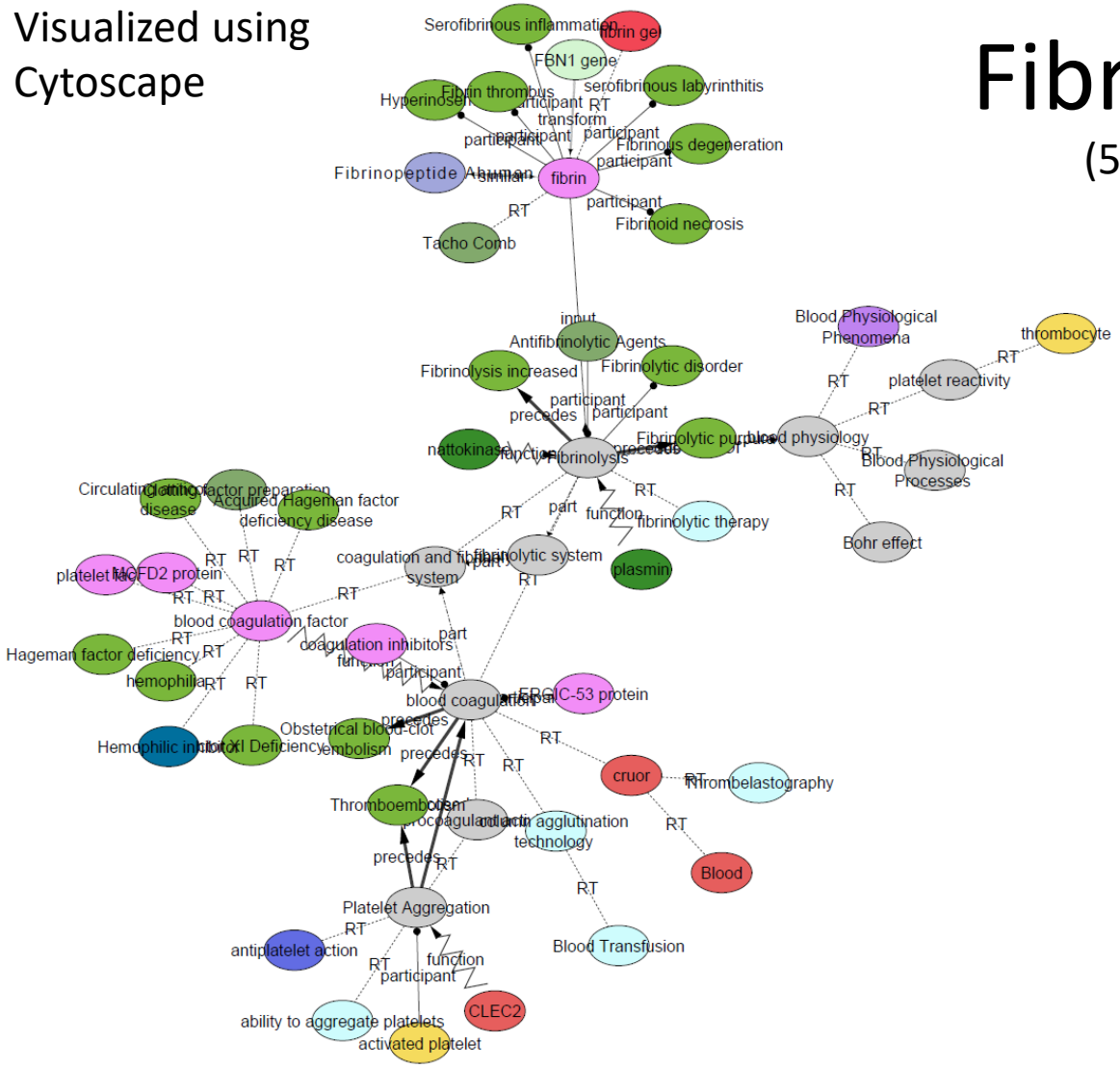
Performing SPARQL queries



Knowledge graph

- Fibrinolysis network
- Bone metabolism turnover (BMT) network
- etc.

Visualized using Cytoscape



Fibrinolysis network

(54 concepts, 57 relations)

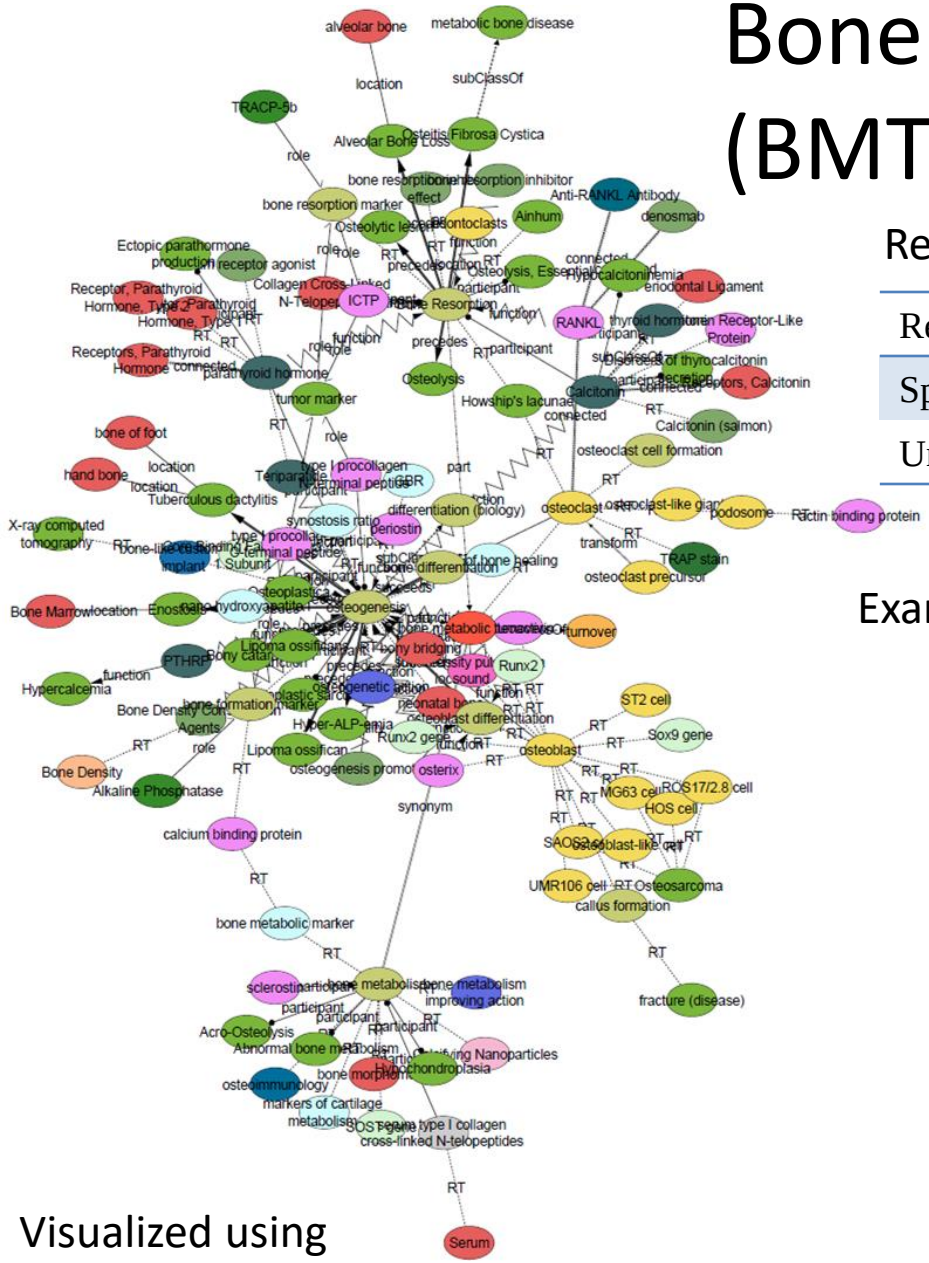
Examples of subjects of concepts in Fibrinolysis network

Subject	Number of concepts
Diseases, pathology, and symptoms	16
Hematology	12
Protein and peptide	7
Medicine and medical cares	5
Drugs	3
Tissue and organ	3
Cytology	2
Enzyme	2
Genetics	2
Immunology	1
Organic compounds	1
Pharmacology	1

Results of RT sub-classification in Fibrinolysis network

Relations	Number of relations
Specified relations (e.g., has function)	28
Unspecified relations (RTs)	29

Bone Metabolic Turnover (BMT) network (139 concepts, 140 relations)



Visualized using Cytoscape

Results of RT sub-classification in BMT network

Relations	Number of relations
Specified relations (e.g., has function)	78
Unspecified relations (RTs)	62

Examples of subjects of concepts in the BMT network

Subject	Number of concepts
Diseases, pathology, and symptoms	28
Tissue and organ	20
Protein and peptide	17
Cytology	15
Development and growth	14
Drugs	7
Medicine and medical cares	7
Genetics	6
Hormone	5
Immunology	4
Enzyme	3
Pharmacology	2

Classification of diseases connected to **fibrinolysis** within three steps (relations)

Group	Concept	Relation	Concept	Relation	Concept	Relation	Concept
F1	fibrinolysis	*precedes*	Fibrinolysis increased				
			Fibrinolytic purpura				
F2		*is participant in*	Fibrinolytic disorder				
F3		*has input*	fibrin	*is participant in*	Fibrin thrombus		
					Fibrinous degeneration		
					Hyperinosemia		
					Fibrinoid Necrosis		
					Serofibrinous inflammation		
					serofibrinous labyrinthitis		
F4		*RT (sibling?)*	blood coagulation	*precedes*	Obstetrical blood-clot embolism		
		is part of	coagulation and fibrinolytic system	*RT*	blood coagulation factor	*RT*	Factor XI Deficiency
		RT (sibling?)	blood coagulation	*succeeds*	platelet aggregation	*precedes*	Thromboembolism
				is function of	blood coagulation factor	*RT*	Hageman factor deficiency
							hemophilia
							Acquired Hageman factor deficiency disease
							Circulating anticoagulant disease

Classification of diseases connected to **BMT** within three steps (relations)

Group	Concept	Relation	Concept	Relation	Concept	Relation	Concept
B1	bone metabolic turnover (BMT)	*has part*	Bone Resorption	*is participant in*	Osteolysis, Essential		
B2				*precedes*	osteolysis		
					Osteolytic lesion		
					Alveolar Bone Loss		
					Osteitis Fibrosa Cystica		
B3				*has participant*	Calcitonin	*is participant in*	Disorders of thyrocalcitonin secretion
B4				*is function of*	parathyroid hormone		Ectopic parathormone production
B5			osteogenesis	*is participant in*	Osteoplastic sarcoma		
B6				*precedes*	Lipoma ossificans		
					Bony cataract		
					Osteoplastica		
					Enostosis		
					tuberculous dactylitis		
B7				*is function of*	PTHRP	*has function*	Hypercalcemia
B8					Calcitonin	*is participant in*	Disorders of thyrocalcitonin secretion
							Hypocalcitoninemia
B9				*is participant in*	parathyroid hormone		Ectopic parathormone production
B10		*synonym*	bone metabolism		Abnormal bone metabolism		
					Acro-Osteolysis		

Extending the refined JST thesaurus using other ontologies, thesauri, and LOD

No. of JST thesaurus concepts having links to concepts of other ontologies in Fibrinolysis, and BMT networks

Ontologies	Fibrinolysis network (54)*	BMT network (139)*
MeSH	21 [38.9%]	56 [40.3%]
ChEBI	2 [3.7%]	4 [2.9%]
Gene Ontology	3 [5.6%]	9 [6.5%]
MEDDRA	7 [13.0%]	11 [7.9%]
SNOMEDCT	20 [37.0%]	36 [25.9%]
NCIT	10 [18.5%]	53 [38.1%]
NDFRT	6 [11.1%]	20 [14.3%]
ICD10	1 [1.9%]	2 [1.4%]
Protein ontology	1 [1.9%]	9 [6.5%]
ICD9	0 [0.0%]	2 [1.4%]
OMIM	4 [7.4%]	16 [11.5%]
FMA	1 [1.9%]	6 [4.3%]

*: The numbers in parentheses donates the numbers of concepts

JST thesaurus concepts corresponding to Gene Ontology (GO) terms in the Fibrinolysis network, with the number of the related human gene products

JST thesaurus concepts in Fibrinolysis network			Gene Ontology (GO) terms		
Label of JST thesaurus concepts	No. of gene products		Label of GO terms	No. of gene products	
Fibrinolysis	2	**9**	fibrinolysis	27	**422**
blood coagulation	0		blood coagulation	338	
platelet aggregation	0		platelet aggregation	57	
others	7		-	-	

Conclusion

- We can handle information regarding relationships among the concepts of various levels and categories by using thirty-five sub-classified relations having meaning.
- Using the sub-classified relations, we classified the sixteen fibrinolysis-related diseases into four groups.
- We can integrate the JST thesaurus with other ontologies, and thesauri such as Gene Ontology by matching the labels and synonyms.
- For example, we extended the fibrinolysis network with 422 human gene products derived from Gene Ontology.
- We consider that it would be a benefit and a contribution for the semantic web research community to provide a knowledge base for reasoning, such as DL reasoning.

Future work

- To update and extend the refined JST thesaurus, and the knowledge graph.
- To carefully design the structure of the relationship between concepts to convert the thesaurus into a more rigorous and solid ontology.
- To consider removing the existing relations in order to avoid duplication and conflict.
- To establish a general methodology for extending/refining an existing thesaurus or ontology.
- To extend the methodology to other science and technology fields.

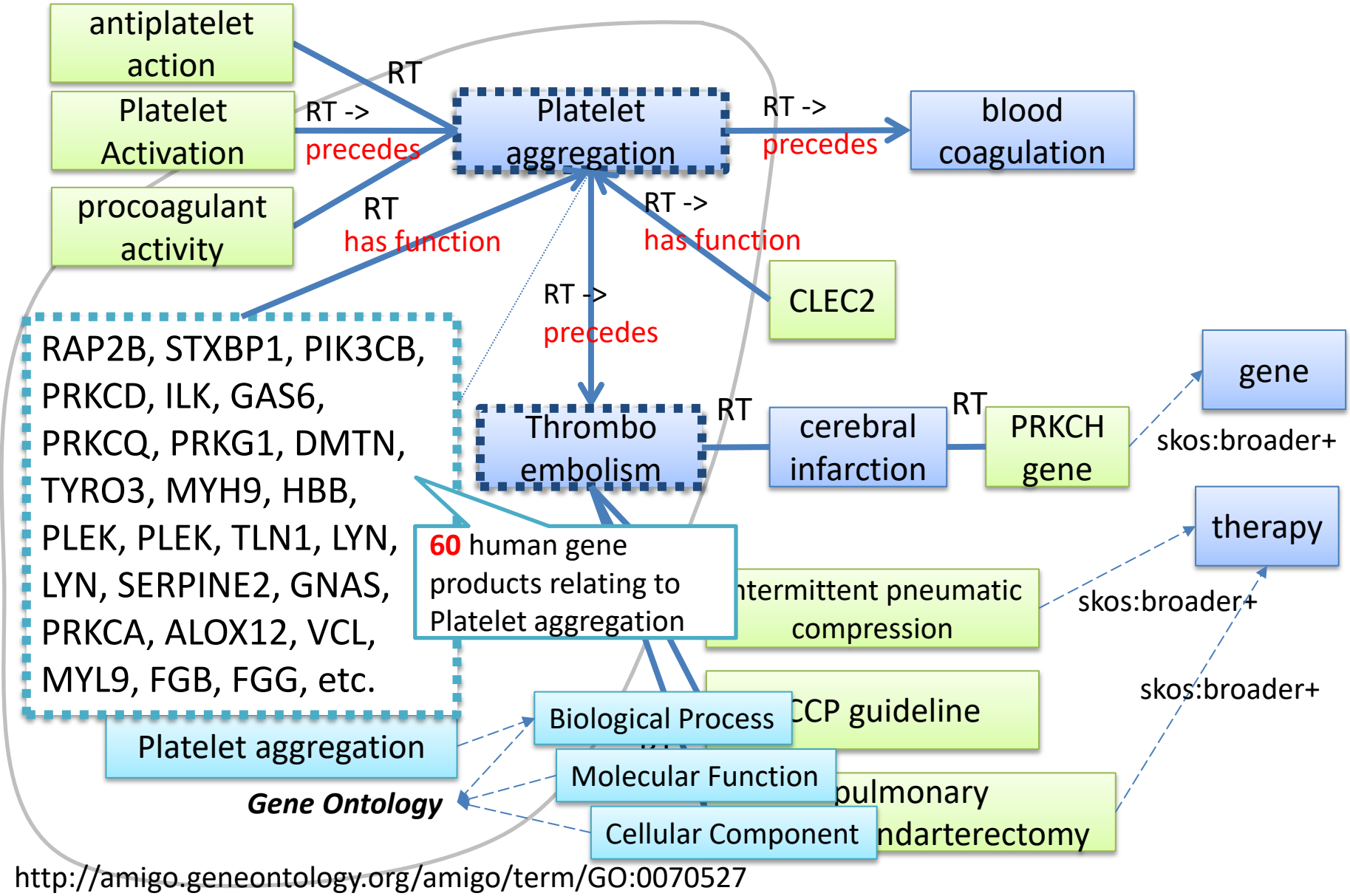
Publication

- At present, the refined (RT sub-classified) JST thesaurus is published by a tentative public SPARQL endpoint with CC BY-NC license (http://lod.hozo.jp/repositories/JstNbcdOnt).

Acknowledgement

- This work was supported by an operating grant from the Japan Science and Technology Agency and JSPS KAKENHI Grant Number JP17H01789.

Thank you for your attention.



http://amigo.geneontology.org/amigo/term/GO:0070527

Mush-up of the information of the refined JST thesaurus and Gene Ontology.

Examples of reasoning using the inheritance approach "is-a," and "part-of composition."

Example 1.

ABC transporter

[*has function*] Biological Transport

[*narrower*] P glycoprotein

Example 2.

spliceosome

[*has part*] splicing factor

splicing factor

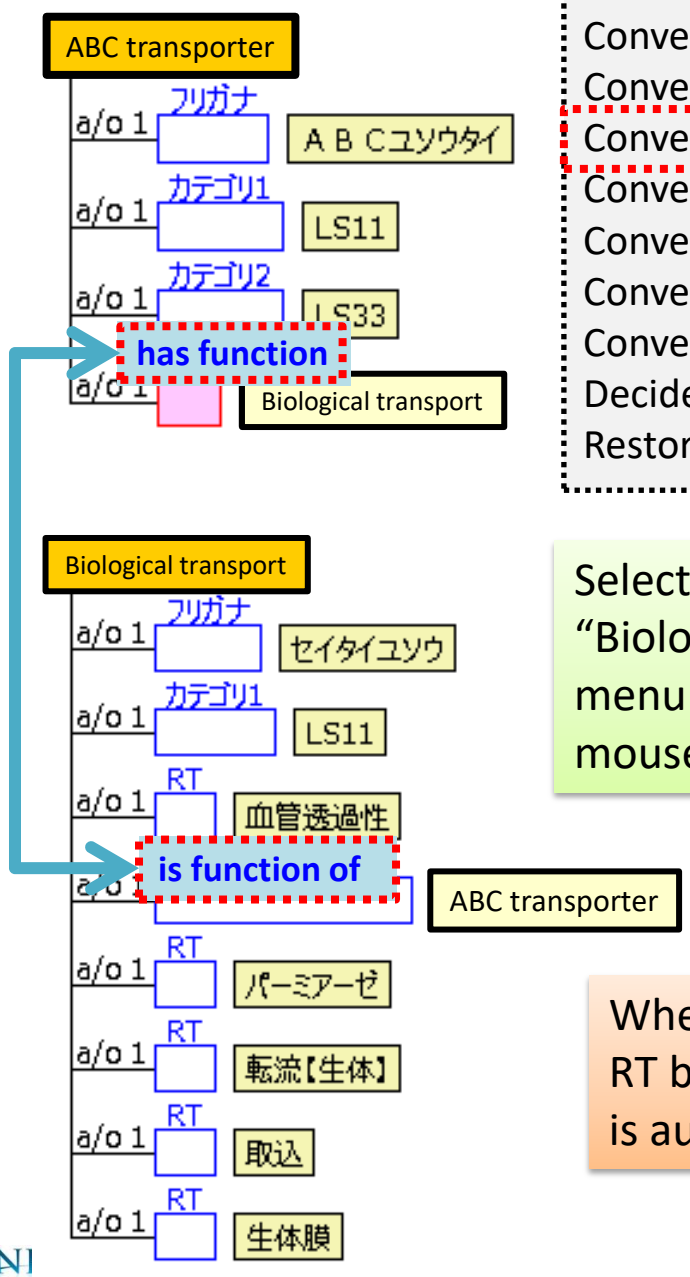
[*has function*] RNA splicing

Related works 1/2

- Examples of the ontological development from thesauri
  - **YAGO** is constructed by unifying the categories and the infoboxes that are automatically extracted from Wikipedia with synsets of WordNet in a rule-based and heuristic method (Suchanek *et al.*, 2007).
  - **AGROVOC** is a thesaurus of agriculture and in the project, it is converted into the ontology by the refining RT in more specific relation, and the modeling using OWL is conducted (Soergel *et al.*, 2004).

Related works 2/2

- Examples of the ontological development by the life-sciences experts and the crowdsourcing
  - Mortensen *et al*. investigated crowdsourcing's performance for validating the relations among concepts in **SNOMED CT** (2015) and **Gene Ontology** (2016).
  - **LEGO** (http://geneontology.org/page/connecting-annotations-lego-models) is the Gene Ontology relating project where modeling semantic relations among biological processes, molecular functions, cellular components, and the related gene products is performed using expert crowdsourcing.



Convert to “BT”
Convert to “NT”
Convert to “has part”
Convert to “is part of”
Convert to “has function”
Convert to “is function of”
Convert to “has attribute”
Convert to “is attribute of”
Convert to “antonym”
Decide “RT”
Restore to “RT”

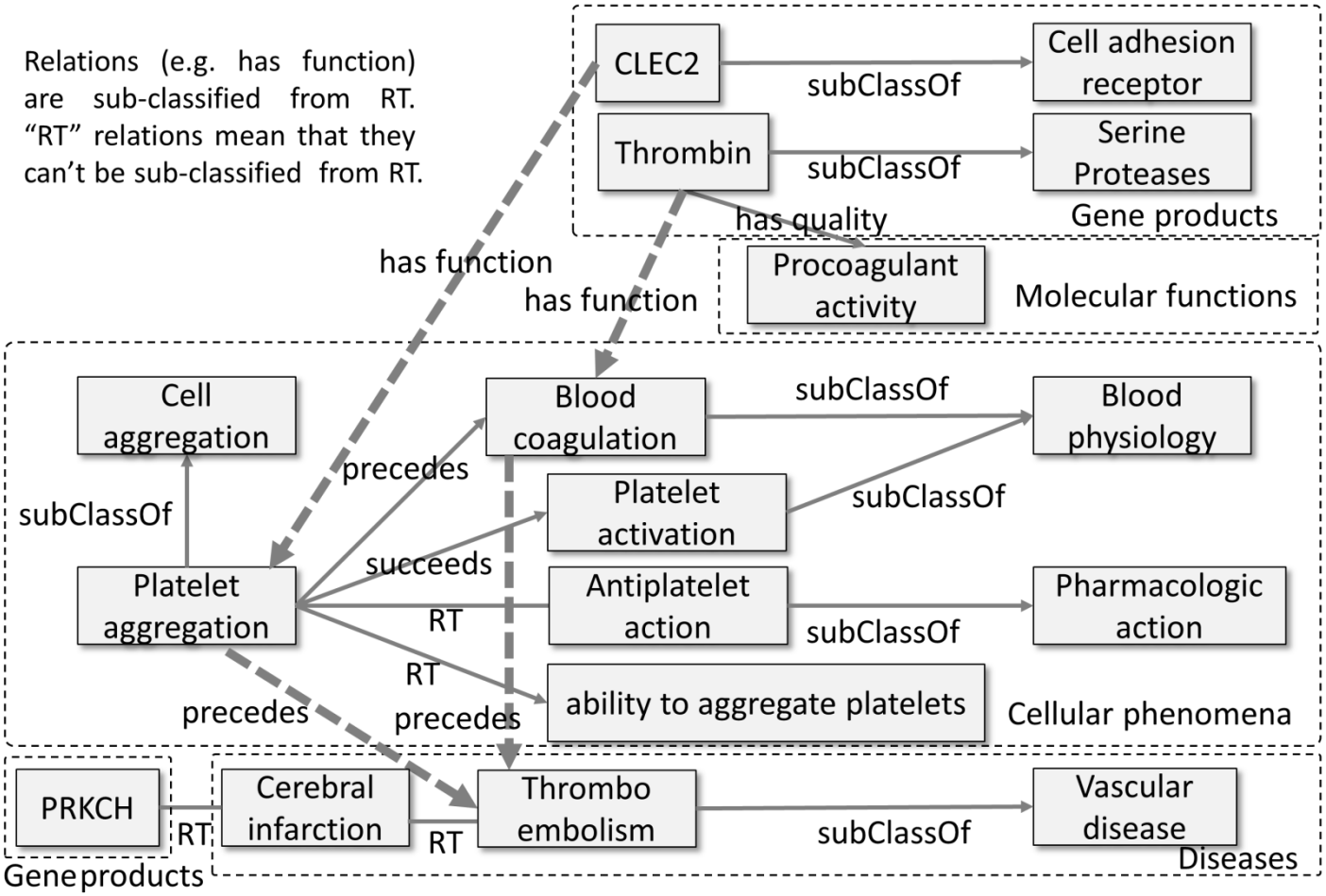
RT sub-classifying using the graphical ontology editor Hozo (in Japanese language)

Ho - zo 法造 an Environment for Building/Using Ontologies www.hozo.jp

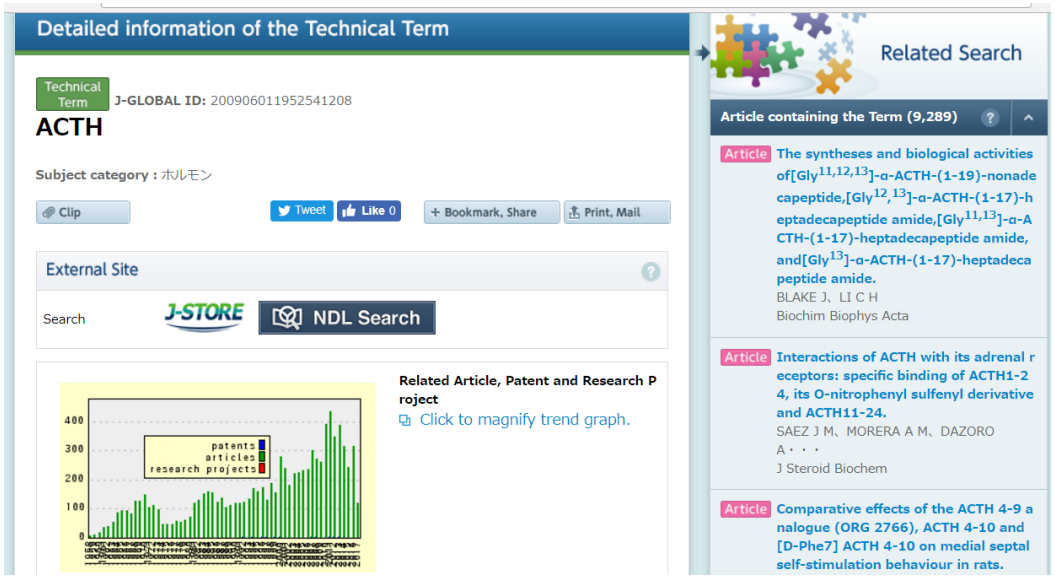
http://www.hozo.jp/

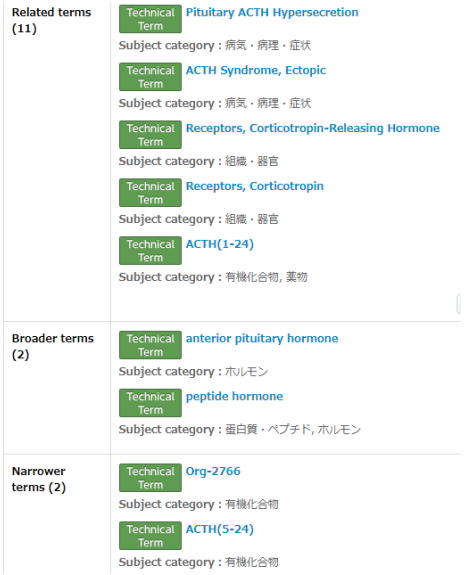
Select the sub-classifying RT between “ABC transporter” and “Biological transport” with mouse left click, open the popup menu with mouse right click, and select “**has function**” on mouse left click in the popup window

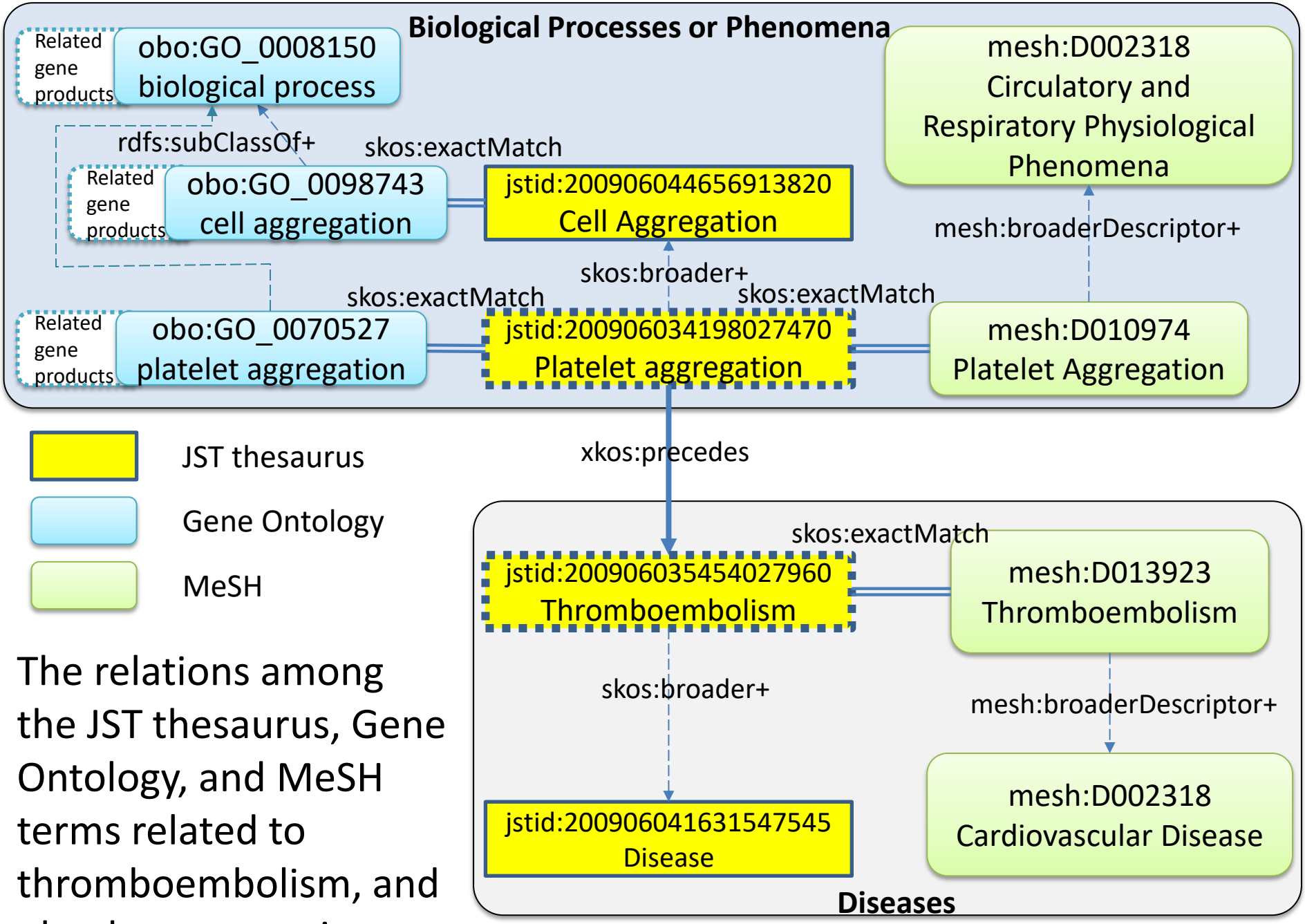
When the RT is converted to “**has function**”, the inverse RT between “Biological transport” and “ABC transporter” is automatically converted to “**is function of**.”


Relations (e.g. has function) are sub-classified from RT. "RT" relations mean that they can't be sub-classified from RT.
CLEC2
subClassOf
Cell adhesion receptor
Thrombin
subClassOf
Serine Proteases
has quality
Gene products
has function
has function
Procoagulant activity
Molecular functions
Cell aggregation
Blood coagulation
subClassOf
Blood physiology
precedes
Platelet activation
subClassOf
subClassOf
succeeds
Platelet aggregation
Antiplatelet action
Pharmacologic action
RT
subClassOf
RT
precedes
precedes
ability to aggregate platelets
Cellular phenomena
PRKCH
RT
Cerebral infarction
RT
Thrombo embolism
subClassOf
Vascular disease
Gene products
Diseases


jglobal.jst.go.jp/en/search/#%7B%22keyword%22%3A%22ACTH%22%2C%22synonym%22%3Anull%2C%22category%22%3A6%2C%22...
J-GLOBAL
科学技術総合リンクセンター
J-GLOBAL home
My J-GLOBAL login
Help
Japan Science and Technology Agency
日本語
About J-GLOBAL
Font size
A
A
ACTH
Technical Term
Thesaurus map
Help
Other names and Synonyms (11)
"Technical Term:ACTH" 's othernames and synonyms(11)
Select All
Clear all
アドレノコルチコトロピン
コルチコトロピン
副じん皮質刺激ホルモン
副腎皮質刺激ホルモン
Adrenocorticotropic Hormone
adrenocorticotrophic hormone
adrenocorticotrophin
adrenocorticotropic hormone
adrenocorticotropin
corticotrophin
corticotropin
Search result of "ACTH"
Add to your search terms to get more hits
ALL 10,089
Researcher 568
Article 9,289
Patent 45
Research Project 4
Organization 0
Technical Term 55
Chemical Subsutance 122
Gene 2
Material 4
Research Resource 0


Detailed information of the Technical Term
Technical Term
J-GLOBAL ID: 200906011952541208
ACTH
Subject category : ホルモン
Clip
Tweet
Like 0
+ Bookmark, Share
Print, Mail
External Site
Search
J-STORE
NDL Search
Related Article, Patent and Research Project
Click to magnify trend graph.
patents
articles
research projects
Related Search
Article containing the Term (9,289)
Article The syntheses and biological activities of[Gly11,12,13]-α-ACTH-(1-19)-nonadecapeptide,[Gly12,13]-α-ACTH-(1-17)-heptadecapeptide amide,[Gly11,13]-α-ACTH-(1-17)-heptadecapeptide amide, and[Gly13]-α-ACTH-(1-17)-heptadecapeptide amide.
BLAKE J、LI C H
Biochim Biophys Acta
Article Interactions of ACTH with its adrenal receptors: specific binding of ACTH1-24, its O-nitrophenyl sulfenyl derivative and ACTH11-24.
SAEZ J M、MORERA A M、DAZORO A・・・
J Steroid Biochem
Article Comparative effects of the ACTH 4-9 analogue (ORG 2766), ACTH 4-10 and [D-Phe7] ACTH 4-10 on medial septal self-stimulation behaviour in rats.

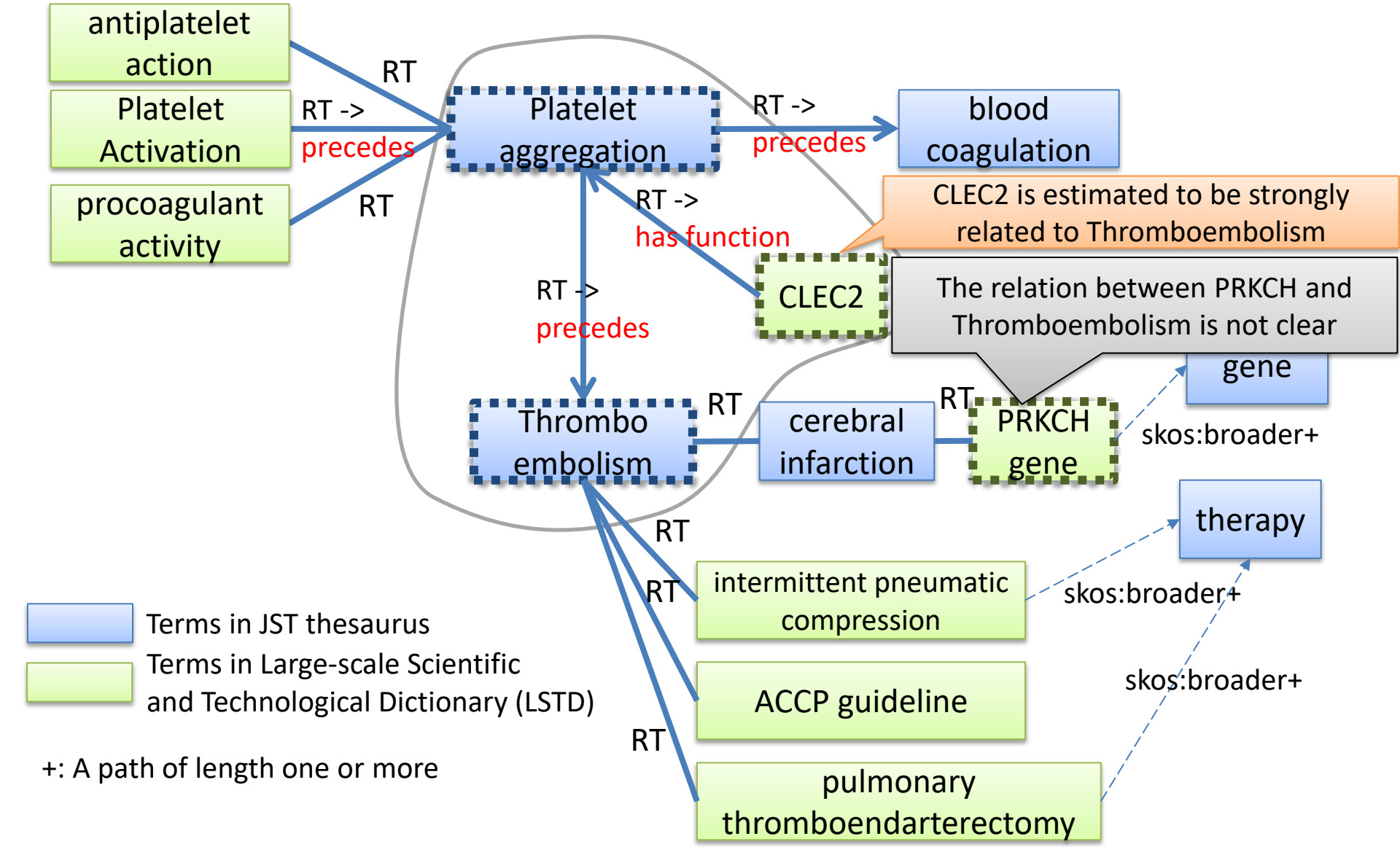

Related terms (11)
Technical Term Pituitary ACTH Hypersecretion
Subject category : 病気・病理・症状
Technical Term ACTH Syndrome, Ectopic
Subject category : 病気・病理・症状
Technical Term Receptors, Corticotropin-Releasing Hormone
Subject category : 組織・器官
Technical Term Receptors, Corticotropin
Subject category : 組織・器官
Technical Term ACTH(1-24)
Subject category : 有機化合物, 薬物
Broader terms (2)
Technical Term anterior pituitary hormone
Subject category : ホルモン
Technical Term peptide hormone
Subject category : 蛋白質・ペプチド, ホルモン
Narrower terms (2)
Technical Term Org-2766
Subject category : 有機化合物
Technical Term ACTH(5-24)
Subject category : 有機化合物



The relations among the JST thesaurus, Gene Ontology, and MeSH terms related to thromboembolism, and platelet aggregation

mesh: <http://id.nlm.nih.gov/mesh/> +: A path of length one or more

Discovering diseases-related gene products from the RT sub-classified JST thesaurus graph



The graph of the relations among terms related to "Thromboembolism" in RT sub-classified JST thesaurus.

Improvement and evaluation of the method of RT sub-classification by using majority decision of a panel of life-sciences experts

An example in RT sub-classification process based on majority decision

e.g. What is relationship between “blood corpuscle” and “blood ”?

	1st candidate	2nd candidate
Curator A	RT	-
Curator B	part Of	-
Curator C	subClassOf	part Of

In 1st candidates, candidate’s relations which the three curators propose are split, and in 1st & 2nd candidates, a relation is agreed by two curators

Biological expert curators propose a relation based on **majority decision**.

part Of

An ontologist confirms the proposed relation (mainly **ontological consistency**)

Correct

Not correct

The manager decides a relation

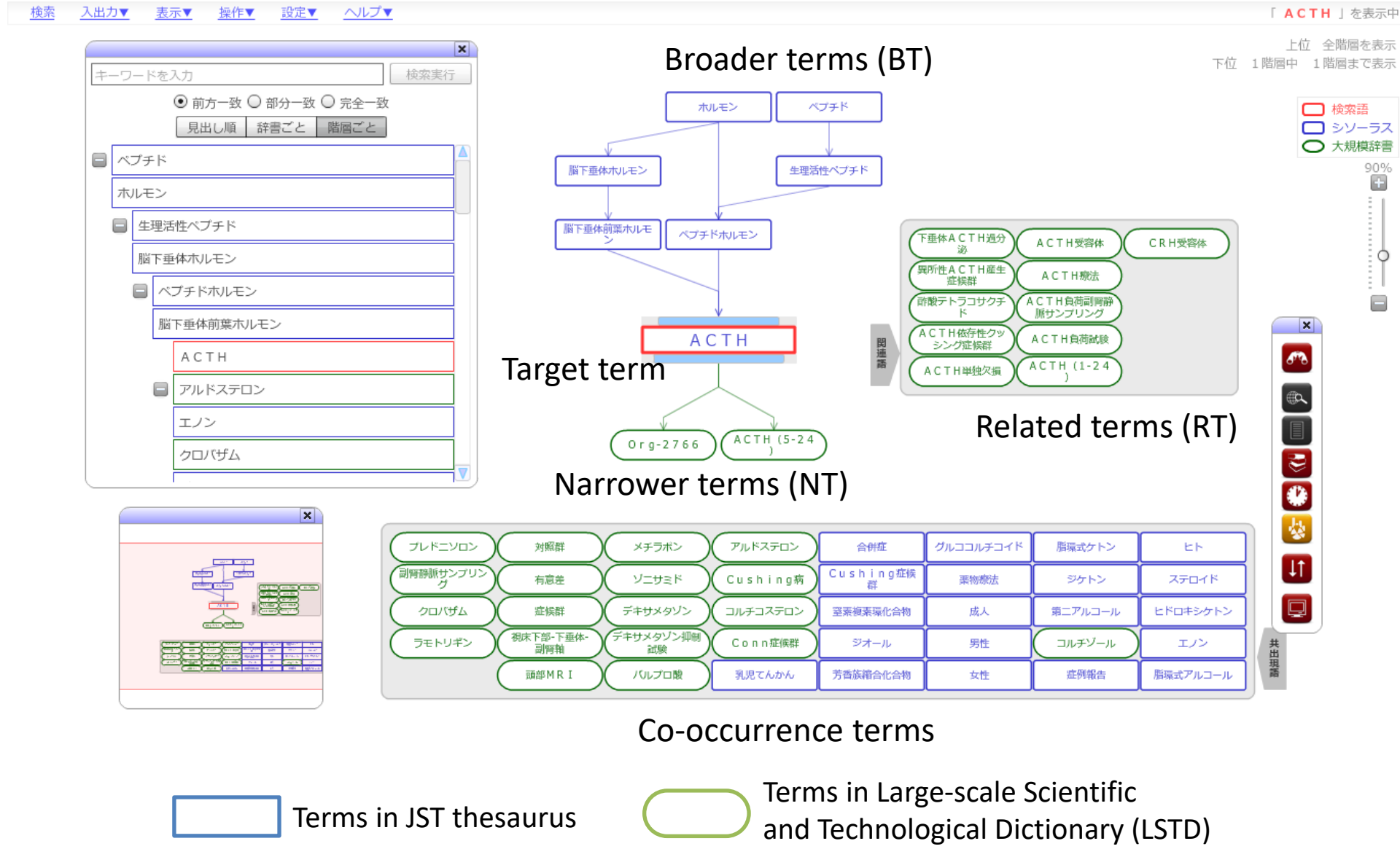
part Of

e.g. subClassOf

Our previous works

- Discovering diseases-related gene products from the RT sub-classified JST thesaurus graph
  - **T. Kushida, et al.,2016**: “Refining JST Thesaurus and Discussing the Effectiveness in Life Science Research” (IESD 2016, co-located with ISWC 2016)

- Improvement and evaluation of the method of RT sub-classification by using majority decision of a panel of life-sciences experts
  - **T. Kushida, et al. 2017**: “Efficient construction of a new ontology for life sciences by subclassifying related terms in the Japan Science and Technology Agency thesaurus” (ICBO 2017)



JST thesaurus map

http://thesaurus-map.jst.go.jp/