Apply SBGN-AF in the LEGO Project

Huaiyu Mi

University of Southern California

Sept. 20, 2013, COMBINE 2013, Paris

Gene Ontology

- The first ontology that was designed as a formal representation of biological knowledge
- Three knowledge domains:
  - molecular function
  - biological process
  - cellular component.

the Gene Ontology

Downloads   Tools   Documentation   Projects   About   Contact

Welcome to the Gene Ontology website!

The Gene Ontology project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. The project provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data from GO Consortium members, as well as tools to access and process this data. Read more about the Gene Ontology…

Search the Gene Ontology Database

Search for genes, proteins or GO terms using AmiGO :

[ ] [GO!]

⦿ gene or protein name   ◯ GO term or ID

AmiGO is the official GO browser and search engine. Browse the Gene Ontology with AmiGO .

Hexokinase

Molecular function: GO:0004396 : hexokinase activity
Biological process: GO:0006096 : glycolysis
Cellular component: GO:0005829 : cytosol

Ashburner, M., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. **25**: p. 25-9.
The Gene Ontology Consortium (2012) The Gene Ontology: enhancements for 2011. *Nucleic Acids Res*. 40:D559

⊡ all : all [377130 gene products]
  ⊞ ▥ GO:0008150 : biological_process [272439 gene products]
    ⊞ ▥ GO:0009987 : cellular process [155181 gene products]
      ⊞ ▥ GO:0044237 : cellular metabolic process [102936 gene products]
        ⊞ ▥ GO:0044262 : cellular carbohydrate metabolic process [5325 gene products]
          ⊟ ▥ **GO:0044275 : cellular carbohydrate catabolic process [1748 gene products]**
            ⊞ ▥ GO:0019405 : alditol catabolic process [66 gene products]
            ⊞ ▥ GO:0046176 : aldonic acid catabolic process [3 gene products]
            ⊡ ▥ GO:0051692 : cellular oligosaccharide catabolic process [4 gene products]
            ⊞ ▥ GO:0044247 : cellular polysaccharide catabolic process [157 gene products]
            ⊞ ▥ GO:0046352 : disaccharide catabolic process [80 gene products]
            ⊡ ▥ GO:0019391 : glucuronoside catabolic process [1 gene product]
            ⊞ ▥ **GO:0006096 : glycolysis [1063 gene products]**
            ⊞ ▥ GO:0016139 : glycoside catabolic process [72 gene products]
            ⊞ ▥ GO:0006098 : pentose-phosphate shunt [367 gene products]
            ⊞ ▣ GO:0043471 : regulation of cellular carbohydrate catabolic process [71 gene products]
            ⊞ ▥ GO:0006258 : UDP-glucose catabolic process [1 gene product]
      ⊞ ▥ GO:0044248 : cellular catabolic process [11927 gene products]
        ⊞ ▥ GO:0044265 : cellular macromolecule catabolic process [6698 gene products]
          ⊟ ▥ **GO:0044275 : cellular carbohydrate catabolic process [1748 gene products]**
            ⊞ ▥ GO:0019405 : alditol catabolic process [66 gene products]
            ⊞ ▥ GO:0046176 : aldonic acid catabolic process [3 gene products]
            ⊡ ▥ GO:0051692 : cellular oligosaccharide catabolic process [4 gene products]
            ⊞ ▥ GO:0044247 : cellular polysaccharide catabolic process [157 gene products]
            ⊞ ▥ GO:0046352 : disaccharide catabolic process [80 gene products]
            ⊡ ▥ GO:0019391 : glucuronoside catabolic process [1 gene product]
            ⊞ ▥ **GO:0006096 : glycolysis [1063 gene products]**
            ⊞ ▥ GO:0016139 : glycoside catabolic process [72 gene products]

GO biological process viewed in AmiGO

What is LEGO Project?

- LEGO is a new development under the Gene Ontology project.
- It is an extension of the existing GO by capturing relationships among various GO terms during the curation of genes.
- Current annotations take a very simple form:
  - Gene_product *implied_relation* GO_class
    - For MF, relation is *has_function*
    - For BP, relation is *involved_in*
    - For CC, relation is *located_in*
- LEGO will enable curators to use the GO to express rich biological statements from the literature
  - Maximize biological knowledge captured by curator
  - Represent complex biology in an accurate, computable manner
  - Prevent
    - "kludges" that use existing terms misleadingly or inconsistently
    - Combinatorial explosion of GO terms

GO annotation 2010

NEDD4

MF
- Ubiquitin-protein ligase activity
- RNA polymerase binding

BP
- Ubiquitin-dependent protein catabolic process
- Cellular response to UV
- Negative regulation of transcription from RNA polymerase II promoter

NEDD4

Fuction

|		
GO:0004842 ubiquitin-protein ligase activity	F	ID
GO:0004842 ubiquitin-protein ligase activity	F	ID
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0016874 ligase activity	F	IE
GO:0019871 sodium channel inhibitor activity	F	IE
GO:0019871 sodium channel inhibitor activity	F	ID
GO:0019904 protein domain specific binding	F	IP
GO:0031698 beta-2 adrenergic receptor binding	F	ID
GO:0043130 ubiquitin binding	F	ID
GO:0050815 phosphoserine binding	F	IE
GO:0050815 phosphoserine binding	F	IS
GO:0050816 phosphothreonine binding	F	IE
GO:0050816 phosphothreonine binding	F	IS
GO:0070063 RNA polymerase binding	F	IP
GO:0070064 proline-rich region binding	F	IE

Process

GO:0000122 negative regulation of transcription from RNA polymerase II promoter	P	IEA	Ensembl Compara
GO:0002250 adaptive immune response	P	IEA	Ensembl Compara
GO:0003151 outflow tract morphogenesis	P	IEA	Ensembl Compara
GO:0003197 endocardial cushion development	P	IEA	Ensembl Compara
GO:0006513 protein monoubiquitination	P	IEA	Ensembl Compara
GO:0006622 protein targeting to lysosome	P	IDA	PMID:17116753
GO:0007041 lysosomal transport	P	IDA	PMID:18544533
GO:0007399 nervous system development	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)
GO:0007528 neuromuscular junction development	P	IEA	Ensembl Compara
GO:0010766 negative regulation of sodium ion transport	P	IEA	Ensembl Compara
GO:0010766 negative regulation of sodium ion transport	P	IDA	PMID:10642508
GO:0010768 negative regulation of transcription from RNA polymerase II promoter in response to UV-induced DNA damage	P	IMP	PMID:17996703
GO:0014068 positive regulation of phosphatidylinositol 3-kinase cascade	P	IMP	PMID:17218260
GO:0016567 protein ubiquitination	P	IEA	Ensembl Compara
GO:0016567 protein ubiquitination	P	IEA	UniPathway2GO
GO:0016567 protein ubiquitination	P	IDA	PMID:18305167
GO:0019048 modulation by virus of host morphology or physiology	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)
GO:0019089 transmission of virus	P	IMP	PMID:15126635
GO:0019221 cytokine-mediated signaling pathway	P	TAS	Reactome:REACT_25229
GO:0030948 negative regulation of vascular endothelial growth factor receptor signaling pathway	P	IEA	Ensembl Compara
GO:0030948 negative regulation of vascular endothelial growth factor receptor signaling pathway	P	ISS	GO_REF:0000024
GO:0031175 neuron projection development	P	IEP	PMID:9990509
GO:0031623 receptor internalization	P	IDA	PMID:18544533

NEDD4

Function		
GO:0004842 ubiquitin-protein ligase activity	F	ID
GO:0004842 ubiquitin-protein ligase activity	F	ID
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0005515 protein binding	F	IP
GO:0016874 ligase activity	F	IE
GO:0019871 sodium channel inhibitor activity	F	IE
GO:0019871 sodium channel inhibitor activity	F	ID
GO:0019904 protein domain specific binding	F	IP
GO:0031698 beta-2 adrenergic receptor binding	F	ID
GO:0043130 ubiquitin binding	F	ID
GO:0050815 phosphoserine binding	F	IE
GO:0050815 phosphoserine binding	F	IS
GO:0050816 phosphothreonine binding	F	IE
GO:0050816 phosphothreonine binding	F	IS
GO:0070063 RNA polymerase binding	F	IP
GO:0070064 proline-rich region binding	F	IE

| Process			
GO:0000122 negative regulation of transcription from RNA polymerase II promoter	P	IEA	Ensembl Compara
GO:0002250 adaptive immune response	P	IEA	Ensembl Compara
GO:0003151 outflow tract morphogenesis	P	IEA	Ensembl Compara
GO:0003197 endocardial cushion development	P	IEA	Ensembl Compara
GO:0006513 protein monoubiquitination	P	IEA	Ensembl Compara
GO:0006622 protein targeting to lysosome	P	IDA	PMID:17116753
GO:0007041 lysosomal transport	P	IDA	PMID:18544533
GO:0007399 nervous system development	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)
GO:0007528 neuromuscular junction development	P	IEA	Ensembl Compara
GO:0010766 negative regulation of sodium ion transport	P	IEA	Ensembl Compara
GO:0010766 negative regulation of sodium ion transport	P	IDA	PMID:10642508
GO:0010768 negative regulation of transcription from RNA polymerase II promoter in response to UV-induced DNA damage	P	IMP	PMID:17996703
GO:0014068 positive regulation of phosphatidylinositol 3-kinase cascade	P	IMP	PMID:17218260
GO:0016567 protein ubiquitination	P	IEA	Ensembl Compara
GO:0016567 protein ubiquitination	P	IEA	UniPathway2GO
GO:0016567 protein ubiquitination	P	IDA	PMID:18305167
GO:0019048 modulation by virus of host morphology or physiology	P	IEA	UniProt Keywords2GO (UniProtKB/Swiss-Prot entries)
GO:0019089 transmission of virus	P	IMP	PMID:15126635
GO:0019221 cytokine-mediated signaling pathway	P	TAS	Reactome:REACT_25229
GO:0030948 negative regulation of vascular endothelial growth factor receptor signaling pathway	P	IEA	Ensembl Compara
GO:0030948 negative regulation of vascular endothelial growth factor receptor signaling pathway	P	ISS	GO_REF:0000024
GO:0031175 neuron projection development	P	IEP	PMID:9990509
GO:0031623 receptor internalization	P	IDA	PMID:18544533

LEGO viewer plugin for Protege
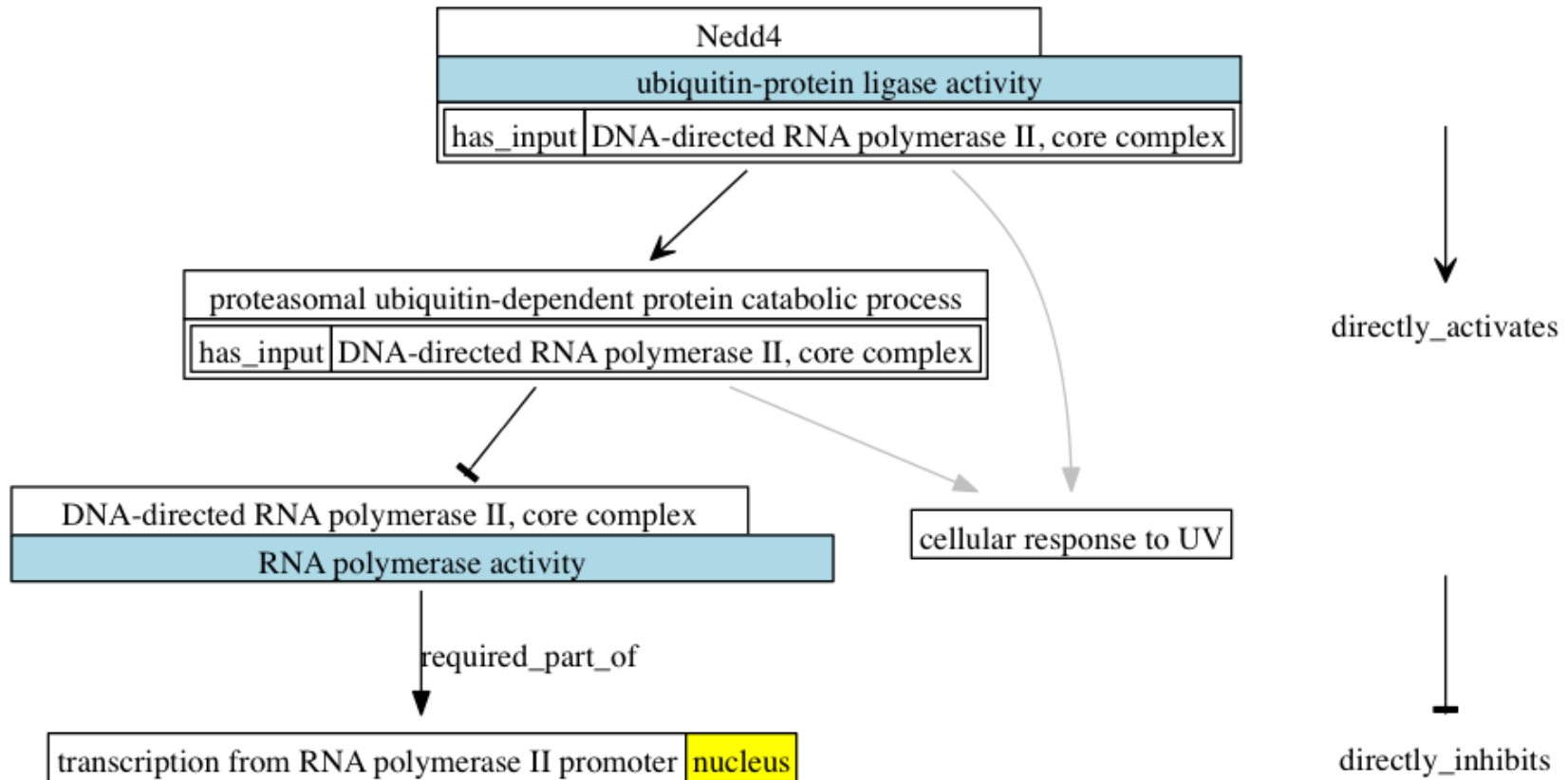
NEDD4 ubiquitin protein ligase activity acting on RNAPII complex involved in positive regulation of proteasomal ubiquitin-dependent catabolic process acting on RNAPII complex involved in negative regulation of RNAPII complex RNA polymerase II activity involved in negative regulation of transcription from RNAPII promoter during cellular response to UV
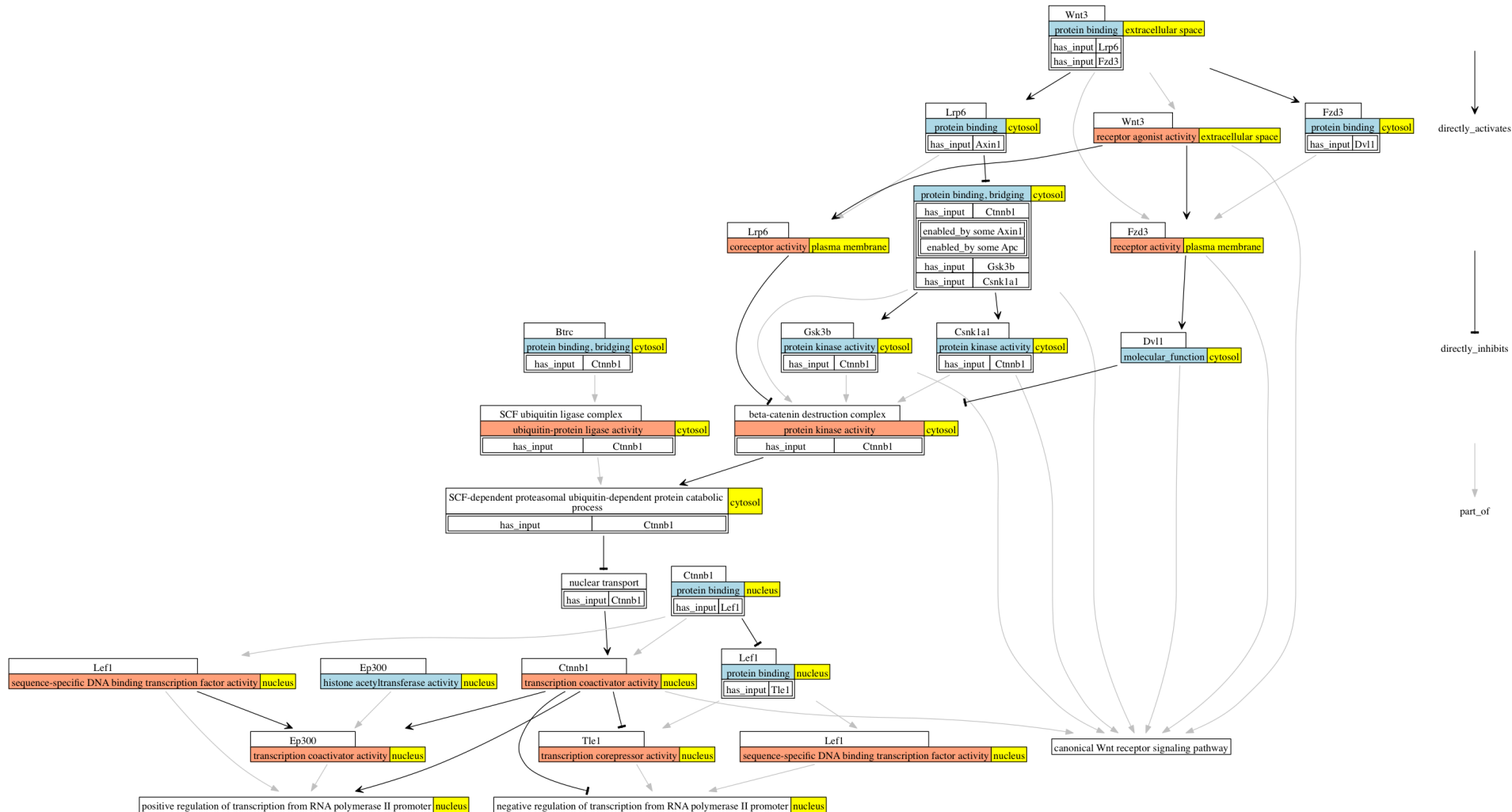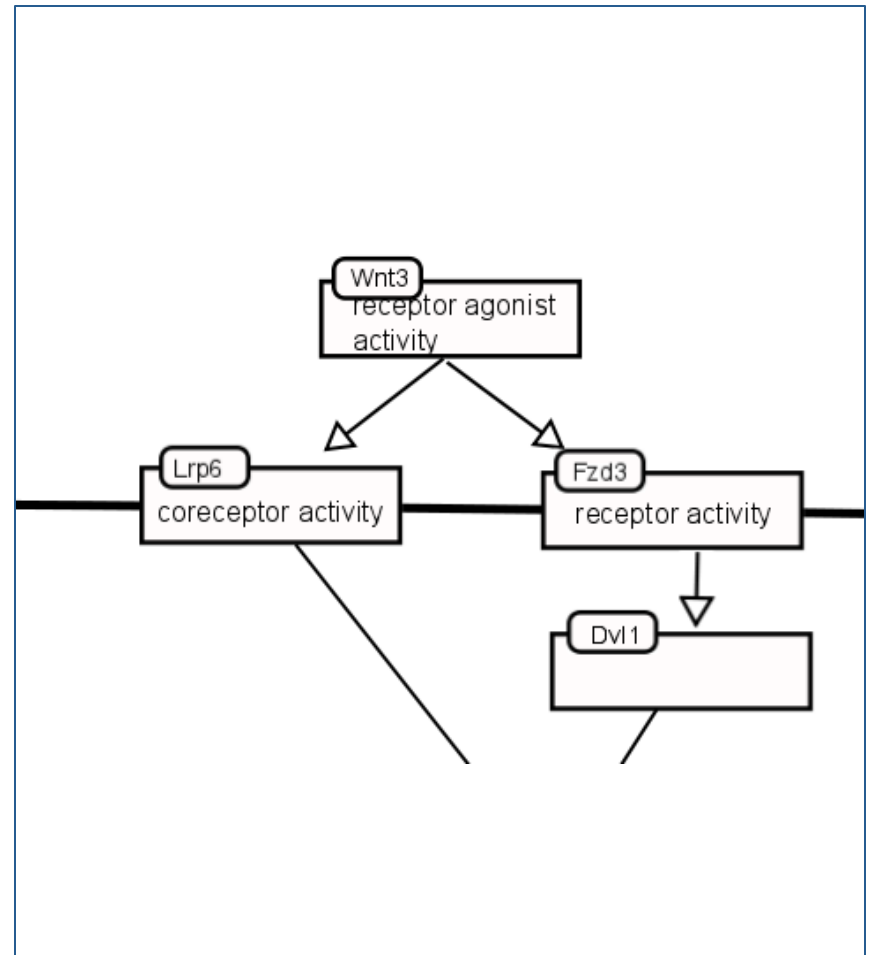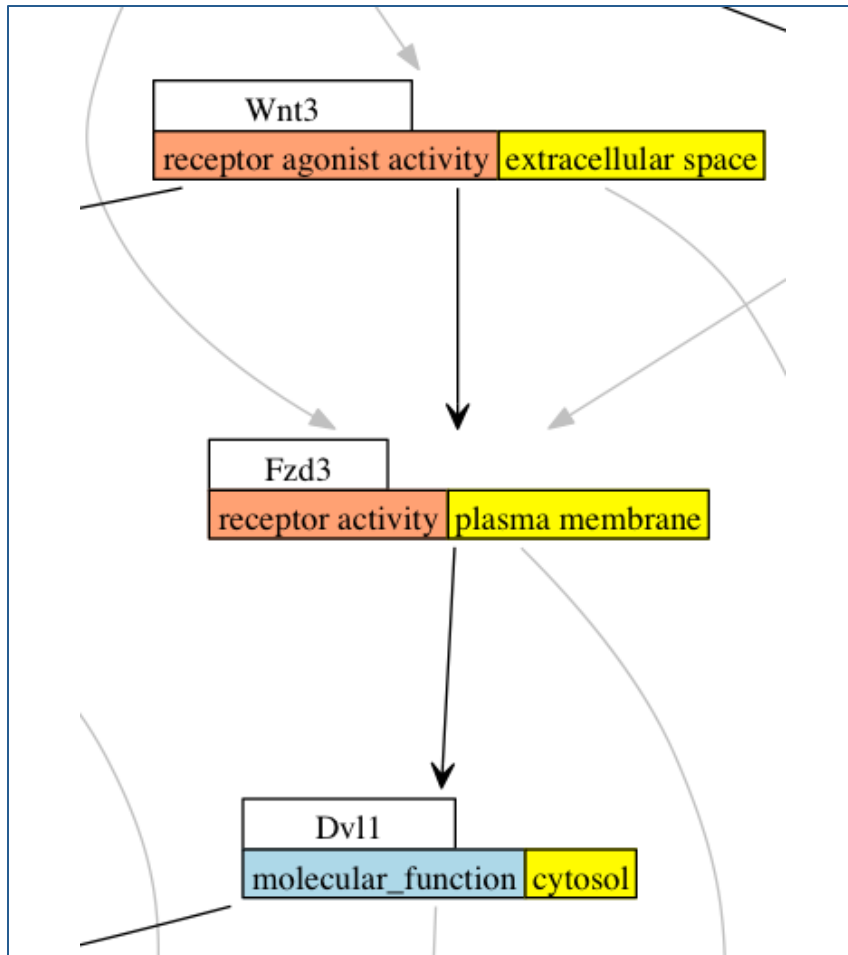
Progress

- Formal spec for construction of LEGO annotation

- Developed plugin for Protégé to enable use as prototype LEGO annotation tool

- Tested with several papers, including review articles on entire pathways
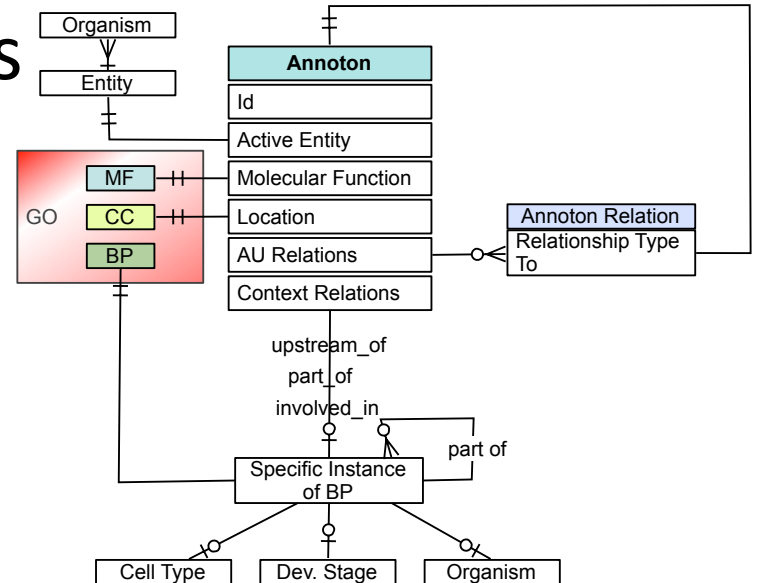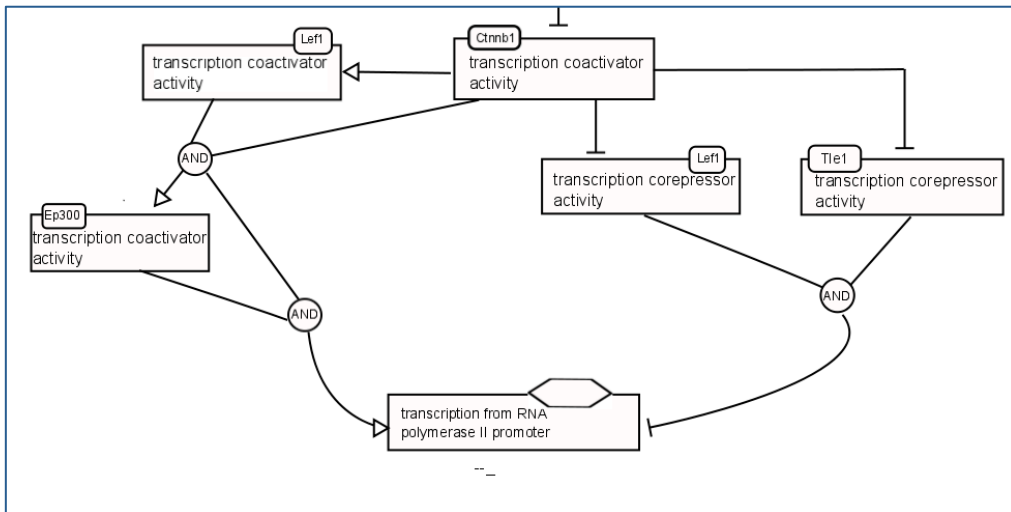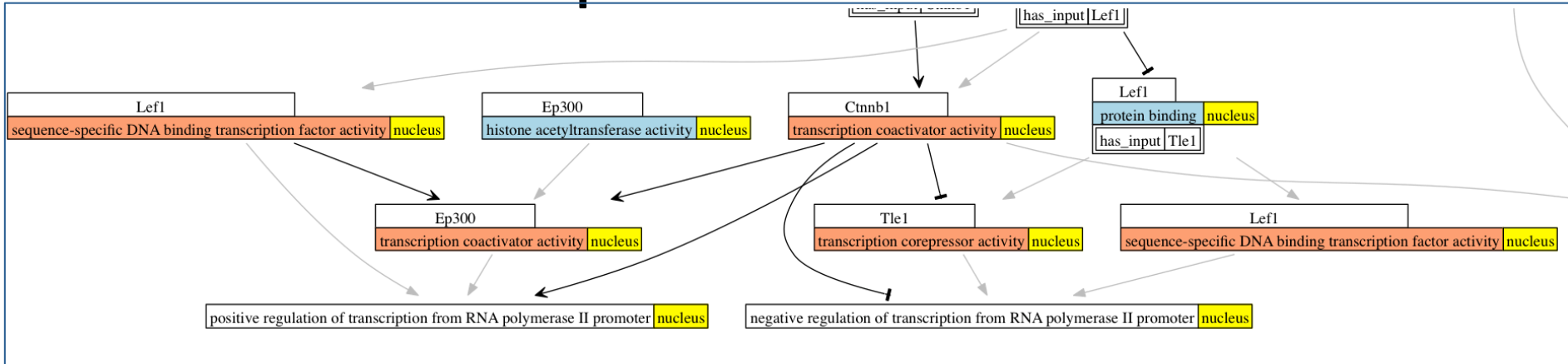
Wnt signaling pathway
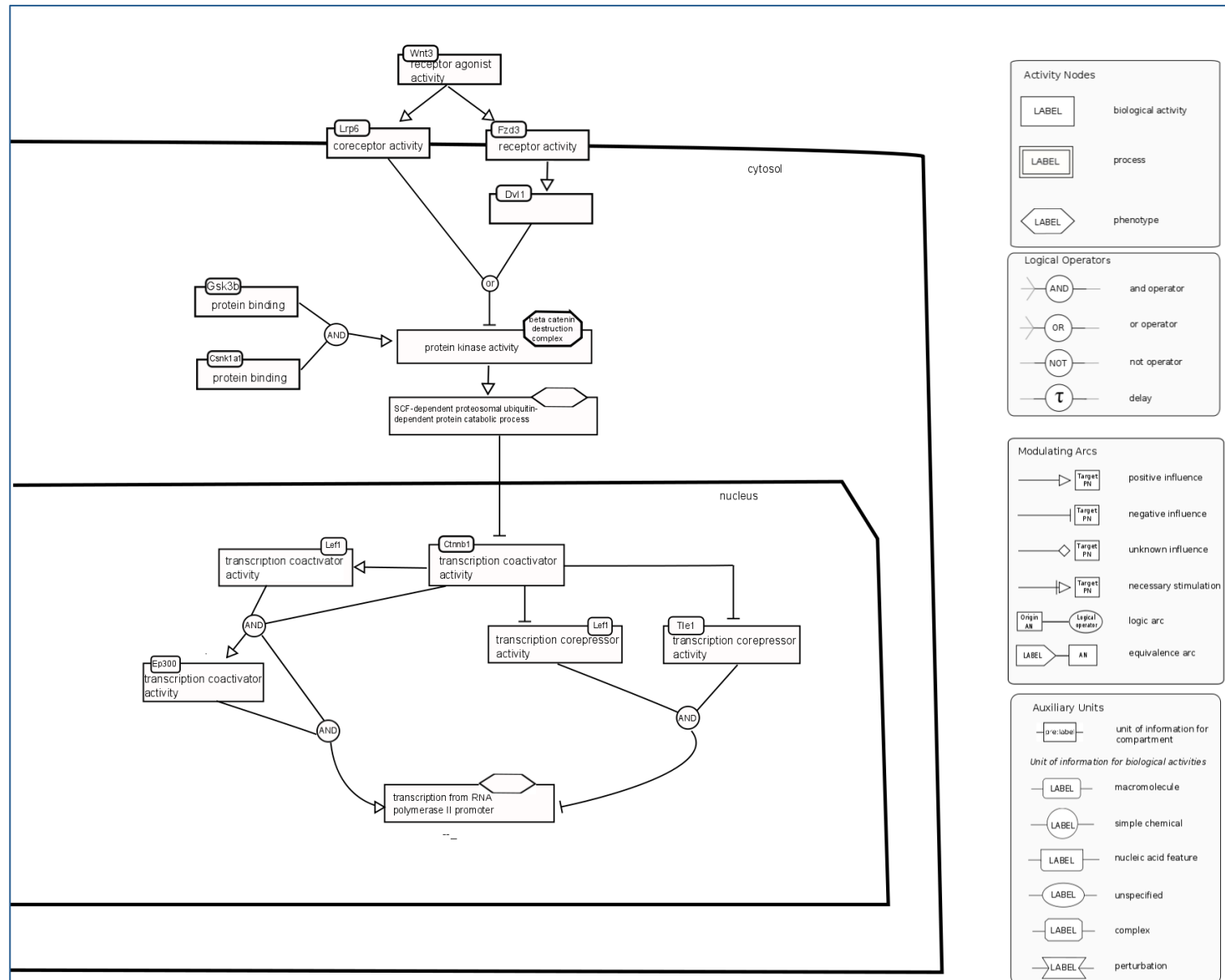
Mapping LEGO to SBGN-AF

Mapping LEGO to SBGN-AF
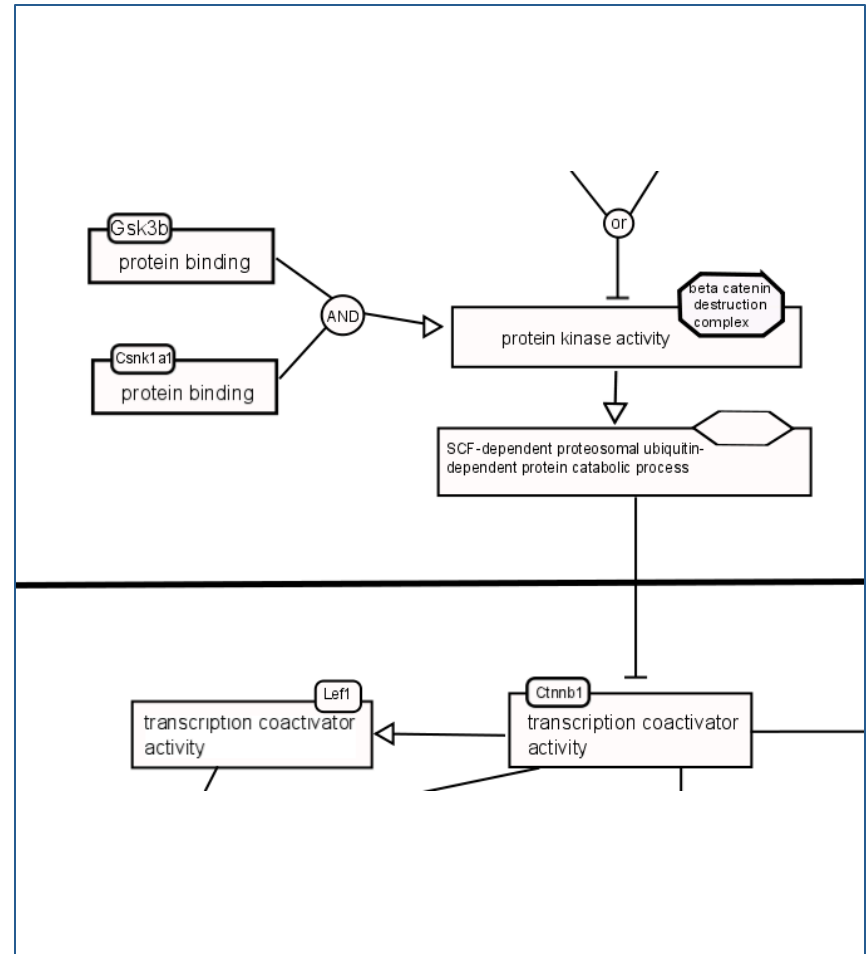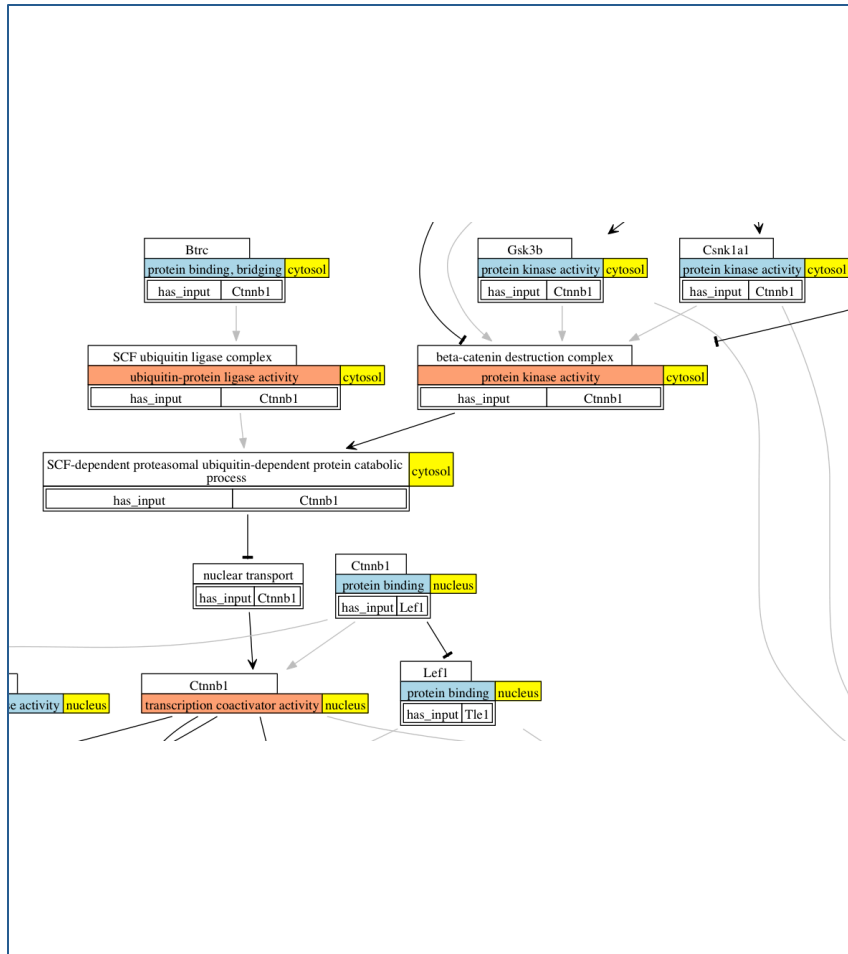
Some part of pathway is easier to represent in PD
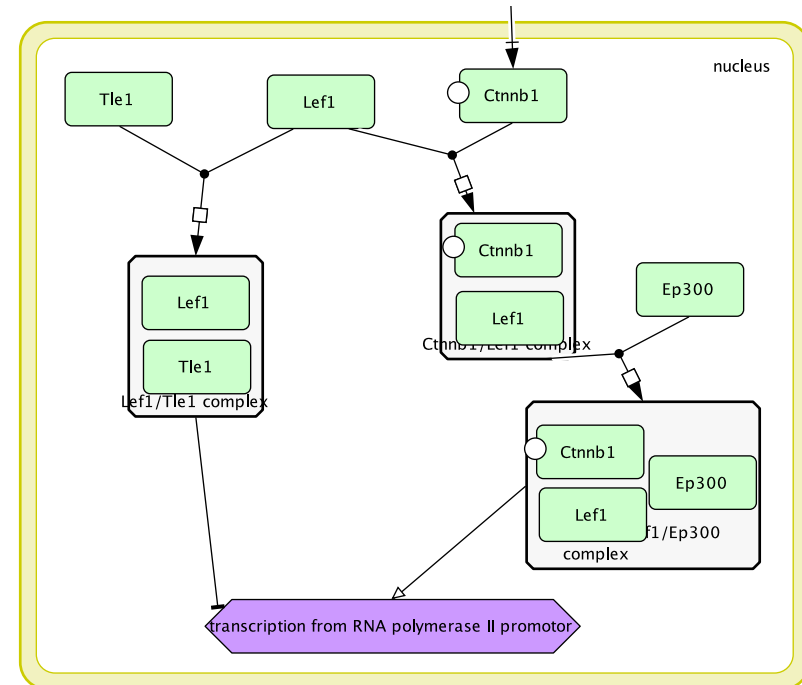
Wnt pathway in SBGN-AF

How SBGN-AF Can Support LEGO

- A LEGO curation tool is being developed.

- SBGN-ML is used to store coordinates.

- LEGO is SBGN-AF in principle, with some exceptions, e.g., entities are used for small chemicals.

- LEGO captures GO specific relationships, such as part_of…

Acknowledgements

GO PIs

- Judy Blake (Jackson Laboratory)
- Paul Thomas (USC)
- Mike Cherry (Stanford)
- Paul Sternberg (Caltech)
- Suzanna Lewis (LBL)

Software development

Lawrence Berkeley Laboratory
- Chris Mungle
- Heiko Dietze

USC
- Anushya Muruganujan

Curation

Jackson Laboratory
- David Hill