Semantic Resources Project

PRO (Protein Ontology) Use Cases for Scientific Communities

Paolo Ciccarese, PhD

Mass General Hospital

Harvard Medical School

Semantic Resources Project

- Collaboration between
  - Mass General Hospital
  - Science Commons
  - Alzforum
- Goal: creating a set of semantic resources
  - Of interest for SWAN and SCF projects
  - Following best practice
- Two examples of PRO related resources:
  - Proteins of interest in AlzSWAN (APP, Tau, Synuclein)
  - Antibodies (immunogens and antigens)

SWAN (Semantic Web Application in Neuromedicine)

- It is a Web-based collaborative program that aims to organize and annotate scientific knowledge about neurodegenerative disorders.
- Collaboration of Mass General Hospital with Alzforum (www.alzforum.org)
- Its goal is to facilitate the formation, development and testing of hypotheses about neurodegenerative diseases.
- It is about scientific discourse

http://swan.mindinformatics.org

http://hypothesis.alzforum.org

AlzSWAN (SWAN for Alzheimer)

SWAN Alzheimer Knowledge Base beta

Semantic Web Applications in Neuromedicine

Log in | SWAN Terms of Use

Home | Statements | Genes-Proteins | Evidence Maps | About

Welcome to the SWAN Alzheimer Knowledge Base

SWAN is the participatory knowledge base of Alzheimer Disease that YOU can help develop. SWAN is all about how you interpret, debate, ask questions and advance the science.

search

» FEATURED CONTRIBUTIONS

The α7 Nicotinic Acetylcholine Receptor gene mediates cognitive deficits and synaptic pathology.
Dziewczapolski Gustavo et al.

Abraham comment on Pollio et al hypothesis "Increased expression of the oligopeptidase THOP1 is a neuroprotective response to Aβ toxicity."
Abraham C R

NMDA receptor hypoactivity (NRHypo) ma have an important contributory role in Alzheimer disease.
Olney J W et al.

click on the title to browse the full content and use the arrows to scroll the list

» HOT TOPICS (browse all hypotheses)

- Aβ accumulation in the brain is the primary event in Alzheimer Disease pathogenesis
- Soluble oligomeric aggregates of Aβ are toxic to neurons and cause AD pathology
- Insoluble fibrillar Aβ leads to AD
- Defective mechanisms of Aβ clearance contribute to AD
- Tau dysfunction mediates neurodegeneration
- ApoE contributes to AD through multiple mechanisms
- Changes in calcium homeostasis may provide a common pathway for the neuropathological changes in AD
- Changes in presenilin function lead to dementia and neurodegeneration in Alzheimer Disease
- Misfolded proteins accumulated into protein aggregates characterizes the pathologic lesions of AD
- The molecular mechanisms of neuronal cell death are involved in the dysfunction and death of neurons in AD
- Synaptic loss appears to be the most powerful and ubiquitous proximate factor leading to the dementia of AD
- Failure of axonal transport might be the underlying basis for neurodegeneration in AD
- Cell membrane properties play a key role in AD Pathophysiology

» MECHANISMS

- Energetics
- Functional Changes of Proteins
- Structural Changes of Proteins

» HOW TO CONTRIBUTE

- BUILD A HYPOTHESIS
- CRITIQUE A HYPOTHESIS
- NOMINATE A KEY PAPER
- HELP FIND CONNECTIONS
- PROPOSE NEW FEATURES
- ADD SUPPORTING EVIDENCE

Contact us!

» KNOWLEDGE BASE

Statements
2137 Research Statements
» 173 Hypotheses
» 48 with Extended annotation
» 125 with Simple annotation
» 1964 Claims
60 Research Questions
45 Comments

Publications
2054 Journal Articles
8 Journal Comments
6 Journal News
33 Web Comments

AlzSWAN Curation Process

Single Molecules of Highly Purified Bacterial Alkaline Phosphatase Have Identical Activity

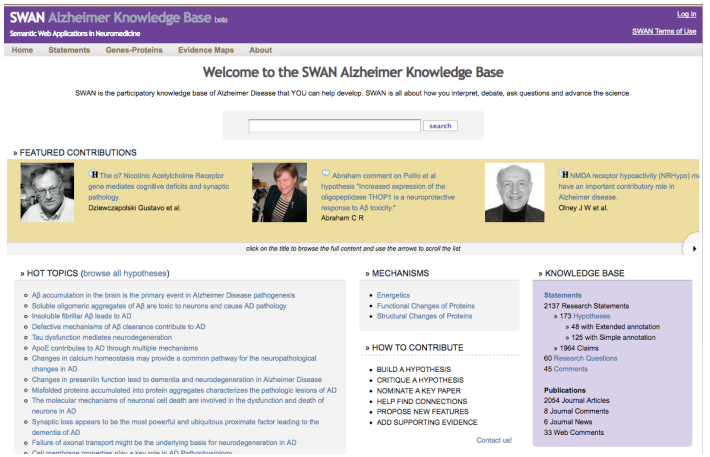
Robert Polakowski, Doug B. Craig,† Alison Skelley, and Norman J. Dovichi*

Abstract: The central paradigm of chemistry is that molecular structure determines molecular function. Details of this paradigm can be tested with single-molecule enzymology, where the activity of individual molecules is studied. In all cases reported thus far, there is a large molecule-to-molecule heterogeneity in activity and activation energy. This heterogeneity must arise from differences in structure. Replicate incubations on the same molecule yield consistent results; the structural heterogeneity must be stable over the time period of the experiment, which can extend over several hours. In this paper, we demonstrate that highly purified molecules of bacterial alkaline phosphatase generate identical activity; structurally identical molecules behave identically. In contrast, the glycosylated mammalian enzyme demonstrates a complex isoelectric focusing pattern and has a dramatic molecule-to-molecule variation in activity and activation energy. Glycosylation affects both the kinetics and energetics of this enzymatically catalyzed reaction.

1. reading

Curator

SWAN: browser

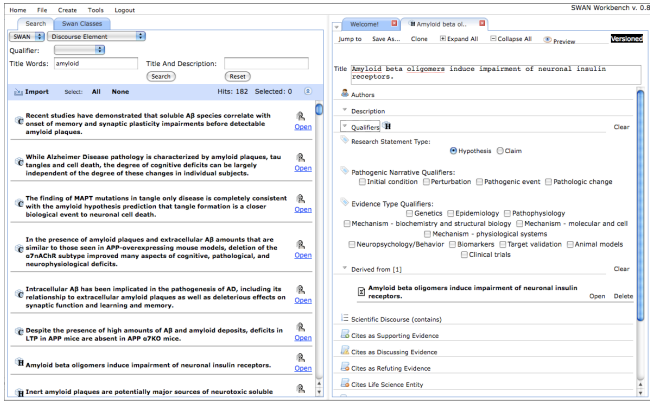


5. publishing

2. extracting

Hypothesis
Discourse List, Citations
Genes, Proteins

3. organizing and encoding

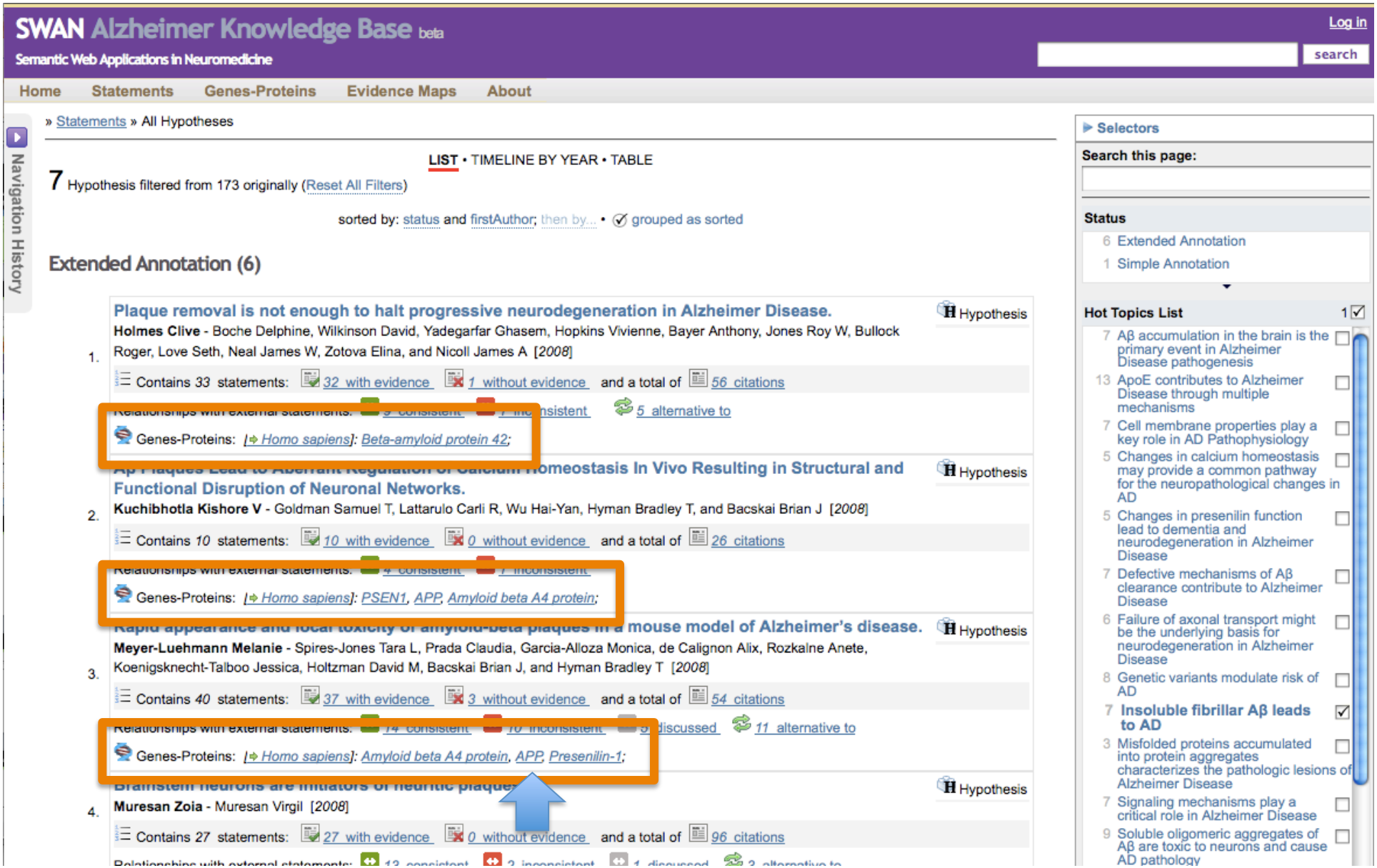


SWAN: workbench

4. connecting

http://tinyurl.com/27sh3ta

Hypothesis list



http://tinyurl.com/2c3lhzn

Hypothesis

Rapid appearance and local toxicity of amyloid-beta plaques in a mouse model of Alzheimer's disease.

3 Comment(s)

Submit a comment Show graph (Experimental!)

Description:

This hypothesis confirms a temporal relationship between plaque deposition and neuritic architecture changes, utilizing data generated from longitudinal in vivo multiphoton microscopy. The authors test and confirm the fundamental tenet of the amyloid hypothesis that amyloid deposition precedes and induces the neuronal abnormalities that underlie dementia, and establish plaques as a critical mediator of neuritic pathology. Plaques form quickly, and lead to the activation and recruitment of microglia as well as increasingly dysmorphic neurites at plaque sites.

Authors: Meyer-Luehmann M Spires-Jones T Prada C Garcia-Alloza M de Calignon A Rozkalne A Koenigsknecht-Talboo J Holtzman D Bacskai B Hyman B

Derived from:

Meyer-Luehmann M, Spires-Jones T, Prada C, Garcia-Alloza M, de Calignon A, Rozkalne A, Koenigsknecht-Talboo J, Holtzman D, Bacskai B, Hyman B
Rapid appearance and local toxicity of amyloid-beta plaques in a mouse model of Alzheimer's disease.
Nature. 2008 Feb 7;451(7179):720-4

Contains 40 Statements:

37 with evidence 3 without evidence and a total of: 54 citations ; Related to external statements: 15 consistent 10 inconsistent 0 discussed 13 with alternatives

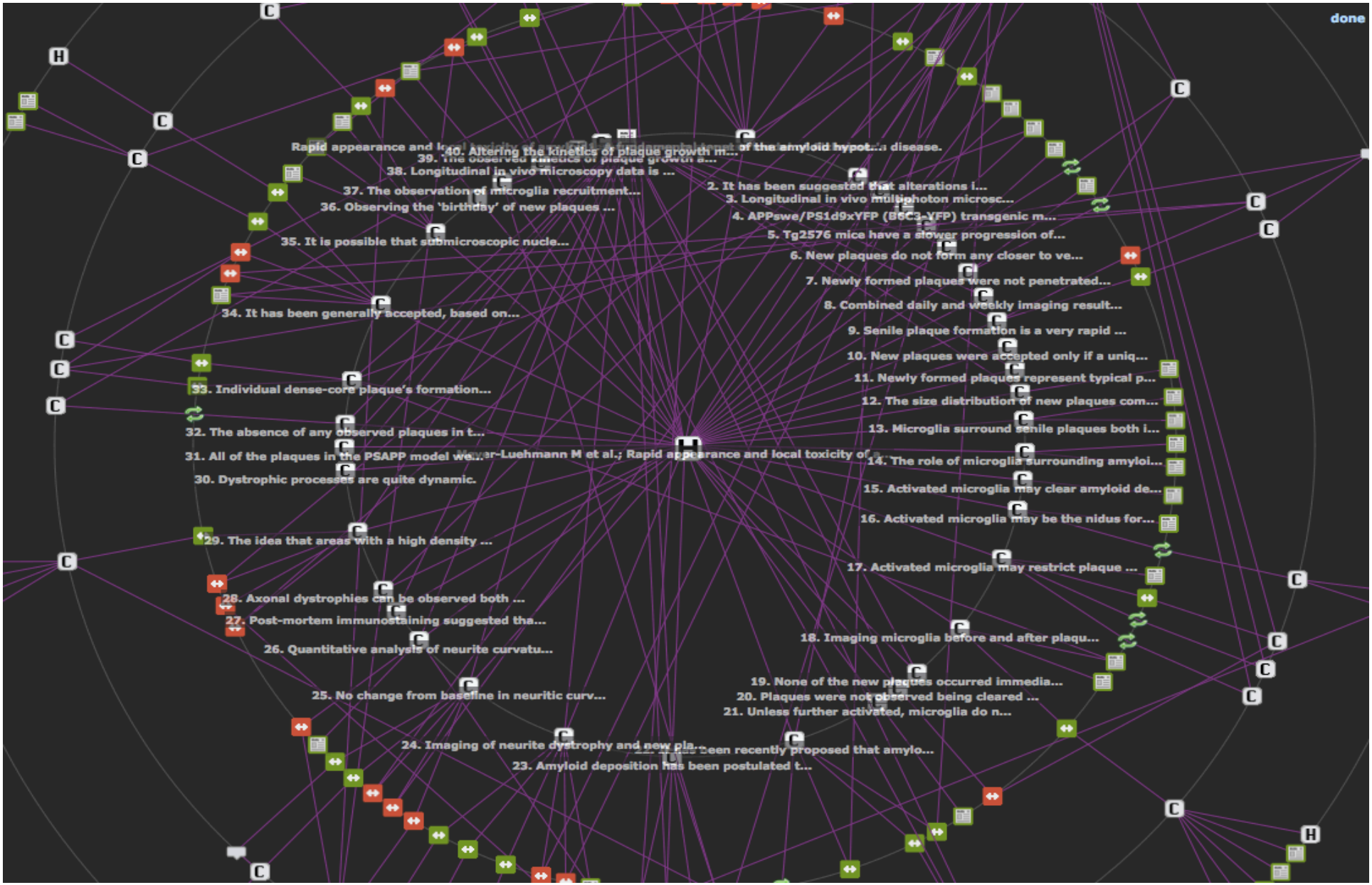
Expand All Details Collapse All Details

1. **A fundamental tenet of the amyloid hypothesis of Alzheimer's disease is that the deposition of amyloid-beta precedes and induces the neuronal abnormalities that underlie dementia.**
   SHOW Details Supporting(1) Consistent(1) Inconsistent(5) Alternative to: (1)
2. **It has been suggested that alterations in axonal trafficking and morphological abnormalities precede and lead to senile plaques**
   SHOW Details Supporting(1)
3. **Longitudinal in vivo multiphoton microscopy has been developed to explore the formation of amyloid plaques and to determine the effects of newly formed dense-cored plaques on the microarchitecture of the brain.**
   SHOW Details Supporting(1) Consistent(1)
4. **APPswe/PS1d9xYFP (B6C3-YFP) transgenic mice begin to show plaque appearance at 5 to 6 months age.**
   SHOW Details Supporting(2) Genes/Proteins: (3)
5. **Tg2576 mice have a slower progression of disease than B6C3-YFP mice, but confirmed the rapidity of plaque development as demonstrated in the B6C3-YFP mice.**
   SHOW Details Supporting(1)
6. **New plaques do not form any closer to vessels than would be expected by chance, in accord with an earlier study of human Alzheimer's disease**
   SHOW Details Supporting(2) Alternative to: (1)

http://tinyurl.com/2c3lhzn

Hypothesis

The hypothesis graph



AlzSWAN proteins management

- When the curators need a new protein they import it automatically from UniProt
- Data are stored according to the SWAN LSEs Ontology (http://swan.mindinformatics.org/spec/1.2/lses.html)
- UniProt does not give access to all the info we need (isoforms, cleavage products…)
- Willing to use PRO ontology and others when available (and demonstrate to cover our needs)

http://swan.mindinformatics.org/spec/1.2/lses.html

AlzSWAN

- In AlzSWAN we have statements related to gene or protein families that do not specify a single species (human vs mouse or rat), but we have also species specific statements
- Protein post-translationally modified have distinct roles:
  - Protein **isoforms** from a single gene have different physiological vs. pathological functions
  - **Cleavage products** have functions distinct from the parent protein
    - Example: APP protein→ Abeta peptides, soluble APP, AICD
  - Proteins are assembled into **protein complexes**, and these are affected by isoform and post-translational modification
    - Example: Tau protein: how many microtubule binding domains (which exons used from gene), phosphorylation state of protein affect MT binding or fibril formation
    - Example: Abeta peptides form dimers, oligomers, amyloid plaque

APP is a gene that encodes a complexity of proteins

- A single locus: single copy per haploid genome
- Multiple splice variant mRNAs
- Mulitiple protein isoforms expressed
  - Some isoforms are ubiquitous, others neuron specific or T cell-specific
- Variety of conserved functional domains throughout length of protein
  - Isoform differences in major domains may have functional consequences
- APP interacts with multiple proteins
- Mutations throughout protein associated with familial Alzheimer Disease
- Cleavage products generated by α-, β- and γ-secretase are involved in both normal and pathogenic protein functions.

SWAN Protein Ontology is NOT just for APP

Ontology term	APP	Tau	Synuclein
Isoforms	10	8	3
Protein Domains	many	MT-binding domains	Tandem repeats
Mutations in disease	AD	FTDP	Parkinson Disease
Heteromulitmeric Protein Complexes	many	Microtubules	many
Homomultimeric Protein Complexes	Oligomers, Plaques	PHF, Tangles Pick bodies	Fibrils, Lewy bodies

Summary of SWAN needs

- Protein Ontology needs vocabulary and relationships that include:
  - **Isoforms**
  - **Splice variants**
  - **Cleavage products**
  - **Mutations**
  - **Protein complexes: homomultimeric and heteromultimeric**
- Tissue-specific expression, protein complexes are important
- Intracellular protein trafficking and post-translational modifications are important

Goals for SWAN:

Research statements attributed to APP, can be specifically attributed to:

- an isoform of APP
- a cleavage product of APP
- a mutated form of APP
- a protein complex formed with APP or it's cleavage products

AlzSWAN curators interaction with PRO

- APP isoforms and cleavage products have been modeled
- These required changes in the PRO model
- Currently working on Protein Complexes

AlzSWAN Knowledge Base

Powered by SWAN (Semantic Web Applications in Neuromedicine)

Log in

AlzSWAN Terms of Use

Home | Statements | Genes-Proteins | Evidence Maps | About

Welcome to AlzSWAN

Where you explore scientific knowledge about Alzheimer disease and share your own ideas, comments and questions in a semantically structured system.

View our tutorials to learn how it works.

search

» FEATURED CONTRIBUTIONS

Preclinical AD associated with progression to symptomatic AD is detected by PET Pittsburgh Compound B (PiB).
Morris John C et al.

Clusterin (ApoJ) and Complement component receptor CR1 genes are associated with Late Onset AD.
Lambert Jean-Charles et al.

The human br
a unique vulnerab
(AD).
Bartzokis George

PRO not only for SWAN!

» HOT TOPICS (browse all hypotheses)

- Biomarkers are useful for predicting and staging Alzheimer disease
- Aβ accumulation in the brain is the primary event in Alzheimer Disease pathogenesis
- Soluble oligomeric aggregates of Aβ are toxic to neurons and cause AD pathology
- Insoluble fibrillar Aβ leads to AD
- Defective mechanisms of Aβ clearance contribute to AD
- Tau dysfunction mediates neurodegeneration
- ApoE contributes to AD through multiple mechanisms
- Changes in calcium homeostasis may provide a common pathway for the neuropathological changes in AD
- Changes in presenilin function lead to dementia and neurodegeneration in Alzheimer Disease
- Misfolded proteins accumulated into protein aggregates characterizes the

» MECHANISMS

- Energetics
- Functional Changes of Proteins
- Structural Changes of Proteins

» HOW TO CONTRIBUTE

- BUILD A HYPOTHESIS
- CRITIQUE A HYPOTHESIS
- NOMINATE A KEY PAPER
- HELP FIND CONNECTIONS
- PROPOSE NEW FEATURES
- ADD SUPPORTING EVIDENCE

Contact us!

» KNOWLEDGE BASE

Statements
2276 Research Statements
» 177 Hypotheses
» 53 with Extended annotation
» 124 with Simple annotation
» 2099 Claims
61 Research Questions
47 Comments

Publications
2277 Journal Articles
8 Journal Comments
6 Journal News

Science Collaboration Framework

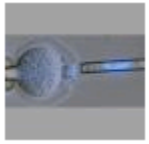
- Reusable software that can be used to develop web-based, collaborative, scientific communities
- Support interdisciplinary scientists in publishing, annotating, sharing and discussing content such as articles, perspectives, interviews and news items, as well as assert personal biographies and research interests
- Applied for the development of the StemBook and PDOnline online communities

http://www.sciencecollaboration.org/

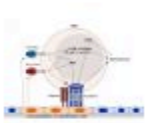
StemBook

HSCI HARVARD STEM CELL INSTITUTE

StemBook is a comprehensive, open-access collection of original, peer-reviewed chapters covering topics related to Stem Cell Biology. Read More

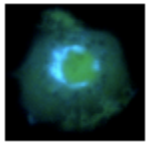

Cellular and nuclear reprogramming

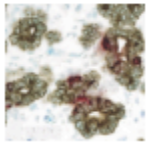

Genomics and proteomics


Renewal

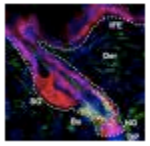

Ectoderm specification and differentiation

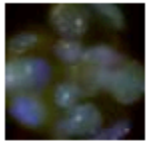

Germ cell and somatic stem cell biology in reproduction

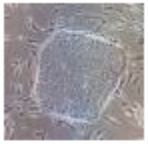

Stem cell immunology

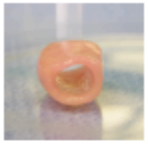

Endoderm specification and differentiation


Mesoderm specification and differentiation


Therapeutic prospects


Epigenetics


Niche biology, homing, and migration


Tissue engineering

New Chapters

The chromatin signature of pluripotent cells
Limbal epithelial stem cells of the cornea
Molecular imaging of stem cells
Stem cells, cancer, and epigenetics

Commentary

Transdifferentiation – Cells Go From Point A to Point B Without "Passing Go"

Published: April 6, 2010

While many stem cell scientists are basking in the afterglow of induced pluripotency, and working on ways to de-differentiate and re-differentiate cells, a few researchers are already thinking about a shortcut.

To read more, click here.

Past Commentaries

Need For Standardization in Stem Cell Research Highlighted in AAAS Webinar

Advances in Imaging Techniques Help Drive Stem Cell Research

pd ONLINE RESEARCH
POWERED BY THE MICHAEL J. FOX FOUNDATION

Welcome Guest | Be a part of PD Online Research

E-mail address: *

Password: *

Log in recover password

Latest Posts PD Guide Funding Opportunities Resources Members News Research Questions
Conference Reports POTW Post here! Reporter Awards LRRK2 Antibodies

Home

search text here

News & Features

MJFF Awards More Than $3 Million to Advance Parkinson's Disease Research

23 Apr 2010

The Michael J. Fox Foundation for Parkinson's Research today announced more than $3 million in previously unannounced awards ending the first quarter of 2010. The funded projects complement the ...

Impax Pharmaceuticals Completes Enrollment in APEX-PD Phase III Trial of IPX066 in Parkinson's Disease

15 Apr 2010

HAYWARD, Calif., Apr 15, 2010 (BUSINESS WIRE) -- Impax Pharmaceuticals, the brand products division of Impax Laboratories, Inc., today announced that it has completed enrollment of its ...

ADAGIO Analysis Demonstrates That the Natural Progression of Clinical Symptoms in Parkinson's Disease May Be Slower in Earlier Stages

19 Apr 2010

JERUSALEM, Israel, Apr 19, 2010 (BUSINESS WIRE) -- --New findings provide further insights into the ADAGIO results which support early treatment initiation with Azilect(R) ...

NPF Awards Nearly $1 Million for Clinical Research

14 Apr 2010

MIAMI—April 14, 2010—The National Parkinson Foundation (NPF) has awarded nearly $1 million dollars to four investigators pursuing clinical research projects in 2010. The research NPF ...

more

Research Questions

Which behavioral outcomes are relevant in rodent models of dyskinesia?

The development of rodent models of LID has improved our ability to conduct studies into the molecular mechanisms of LID as well as to screen new therapeutic strategies. However, it is much harder to detect and analyze rodent behavior relevant to L-DOPA-

Contributions

Most Recent | Editor's Choice

Comment on Pfeiffer and McLaughlin (2010)

By: Lorraine Iacovitti, Thomas Jefferson 23 Apr 2010 11:30 AM EST

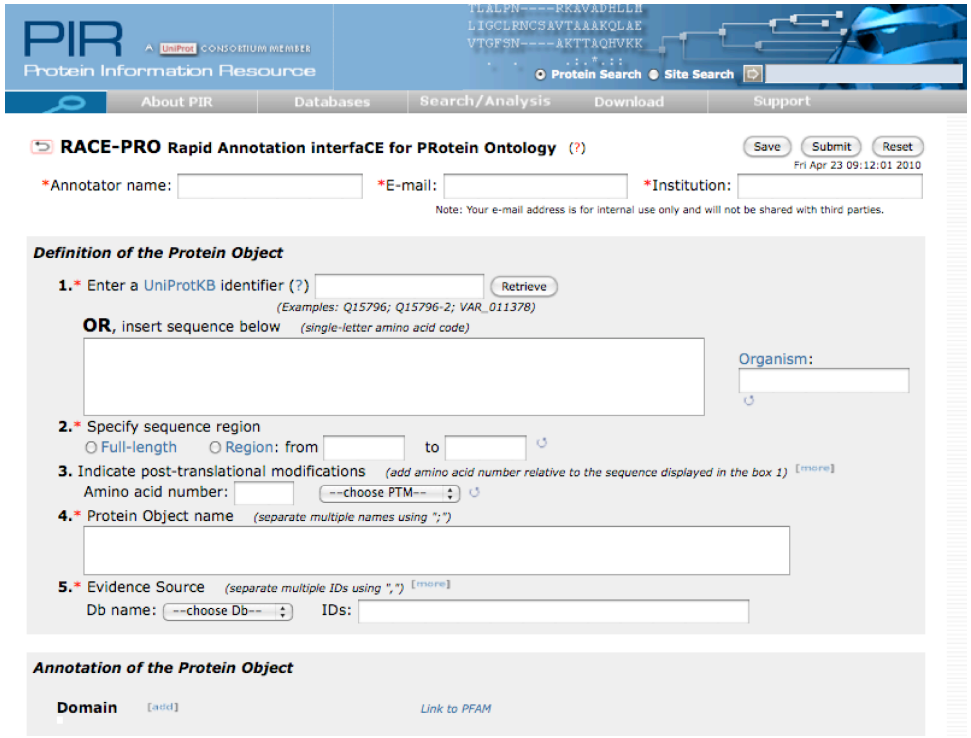
PRO is attractive for our communitites

- AlzSWAN (http://hypothesis.alzforum.org/)
- StemBook (www.stembook.org)
- PDOnline (http://www.pdonlineresearch.org/)
- Others coming up…

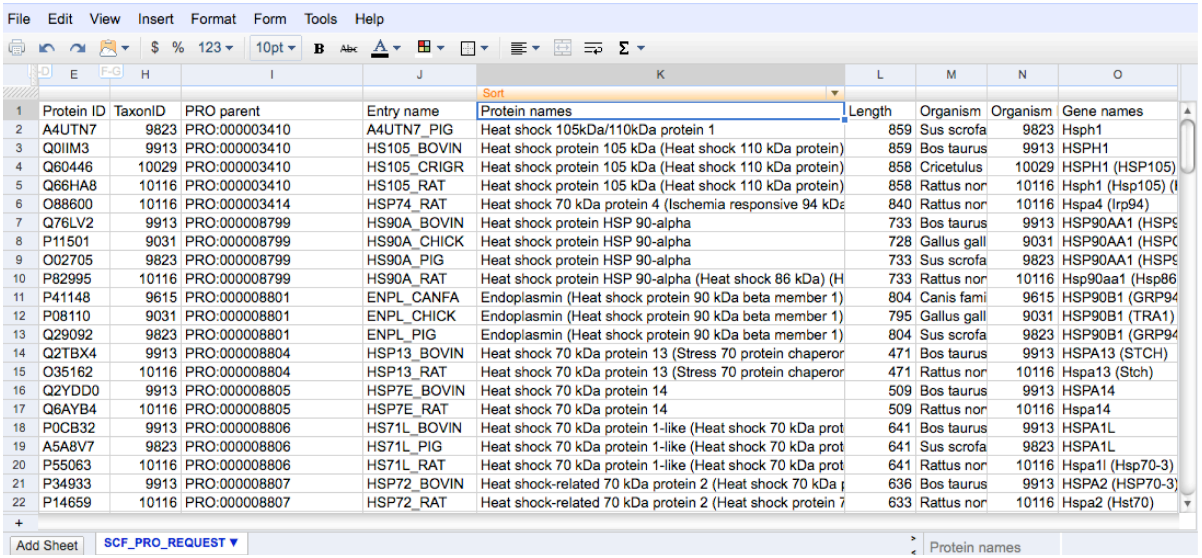
Some thoughts about the PRO terms submission process

PRO terms are submitted by hand

- RACE-PRO web form
  - UniprotKB identifier
  - Annotations
  - Database xrefs
  - “comments”
- Some requests can’t be submitted through RACE-PRO
  - Protein families are submitted through SourceForge



Batch request format is better for integrating with applications



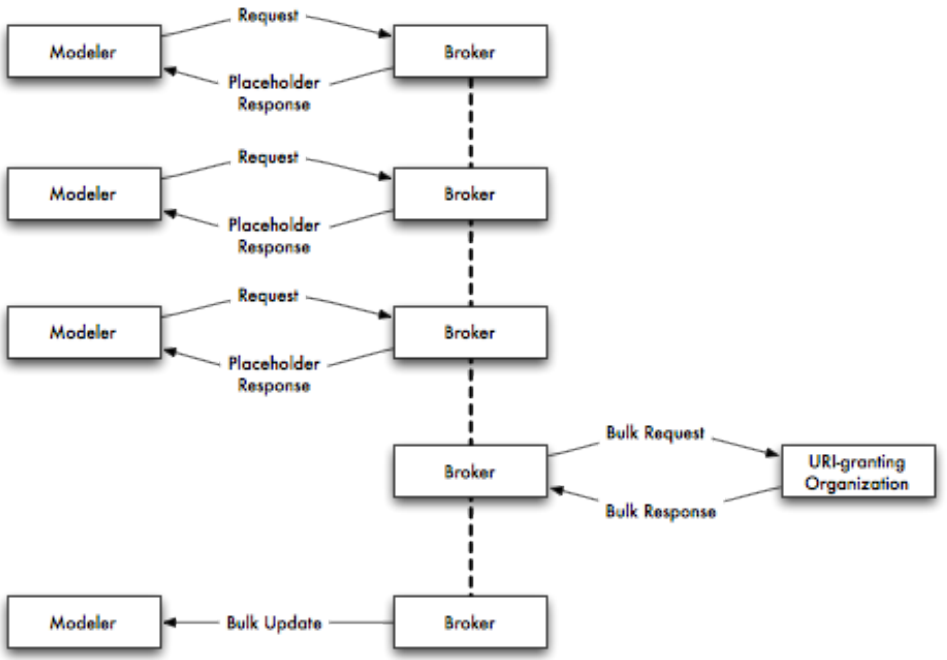
	Protein ID	TaxonID	PRO parent	Entry name	Protein names	Length	Organism	Organism	Gene names
2	A4UTN7	9823	PRO:000003410	A4UTN7_PIG	Heat shock 105kDa/110kDa protein 1	859	Sus scrofa	9823	Hsph1
3	Q0IIM3	9913	PRO:000003410	HS105_BOVIN	Heat shock protein 105 kDa (Heat shock 110 kDa protein)	859	Bos taurus	9913	HSPH1
4	Q60446	10029	PRO:000003410	HS105_CRIGR	Heat shock protein 105 kDa (Heat shock 110 kDa protein)	858	Cricetulus	10029	HSPH1 (HSP105)
5	Q66HA8	10116	PRO:000003410	HS105_RAT	Heat shock protein 105 kDa (Heat shock 110 kDa protein)	858	Rattus nor	10116	Hsph1 (Hsp105) (
6	O88600	10116	PRO:000003414	HSP74_RAT	Heat shock 70 kDa protein 4 (Ischemia responsive 94 kDa	840	Rattus nor	10116	Hspa4 (Irp94)
7	Q76LV2	9913	PRO:000008799	HS90A_BOVIN	Heat shock protein HSP 90-alpha	733	Bos taurus	9913	HSP90AA1 (HSP9
8	P11501	9031	PRO:000008799	HS90A_CHICK	Heat shock protein HSP 90-alpha	728	Gallus gall	9031	HSP90AA1 (HSPC
9	O02705	9823	PRO:000008799	HS90A_PIG	Heat shock protein HSP 90-alpha	733	Sus scrofa	9823	HSP90AA1 (HSP9
10	P82995	10116	PRO:000008799	HS90A_RAT	Heat shock protein HSP 90-alpha (Heat shock 86 kDa) (H	733	Rattus nor	10116	Hsp90aa1 (Hsp86
11	P41148	9615	PRO:000008801	ENPL_CANFA	Endoplasmin (Heat shock protein 90 kDa beta member 1)	804	Canis fami	9615	HSP90B1 (GRP94
12	P08110	9031	PRO:000008801	ENPL_CHICK	Endoplasmin (Heat shock protein 90 kDa beta member 1)	795	Gallus gall	9031	HSP90B1 (TRA1)
13	Q29092	9823	PRO:000008801	ENPL_PIG	Endoplasmin (Heat shock protein 90 kDa beta member 1)	804	Sus scrofa	9823	HSP90B1 (GRP94
14	Q2TBX4	9913	PRO:000008804	HSP13_BOVIN	Heat shock 70 kDa protein 13 (Stress 70 protein chaperon	471	Bos taurus	9913	HSPA13 (STCH)
15	O35162	10116	PRO:000008804	HSP13_RAT	Heat shock 70 kDa protein 13 (Stress 70 protein chaperon	471	Rattus nor	10116	Hspa13 (Stch)
16	Q2YDD0	9913	PRO:000008805	HSP7E_BOVIN	Heat shock 70 kDa protein 14	509	Bos taurus	9913	HSPA14
17	Q6AYB4	10116	PRO:000008805	HSP7E_RAT	Heat shock 70 kDa protein 14	509	Rattus nor	10116	Hspa14
18	P0CB32	9913	PRO:000008806	HS71L_BOVIN	Heat shock 70 kDa protein 1-like (Heat shock 70 kDa prot	641	Bos taurus	9913	HSPA1L
19	A5A8V7	9823	PRO:000008806	HS71L_PIG	Heat shock 70 kDa protein 1-like (Heat shock 70 kDa prot	641	Sus scrofa	9823	HSPA1L
20	P55063	10116	PRO:000008806	HS71L_RAT	Heat shock 70 kDa protein 1-like (Heat shock 70 kDa prot	641	Rattus nor	10116	Hspa1l (Hsp70-3)
21	P34933	9913	PRO:000008807	HSP72_BOVIN	Heat shock-related 70 kDa protein 2 (Heat shock 70 kDa p	636	Bos taurus	9913	HSPA2 (HSP70-3)
22	P14659	10116	PRO:000008807	HSP72_RAT	Heat shock-related 70 kDa protein 2 (Heat shock protein 7	633	Rattus nor	10116	Hspa2 (Hst70)

*Manually curated HSPs related antibodies

- Tabular format
  - Essential fields from RACE-PRO
  - “Tracer ID” tracks individual rows
- Supports automated terms management processes
  - Tracer ID can be used as a “temporary” URI

A “Term Broker” should manage interactions between PRO and applications

- A term broker is server software that assembles batch requests for PRO
  - Receives requests from the modeler
  - Produces placeholder URIs
- Submissions to PRO are made in batch
  - Additional protocol supports substitution of placeholder with final URIs



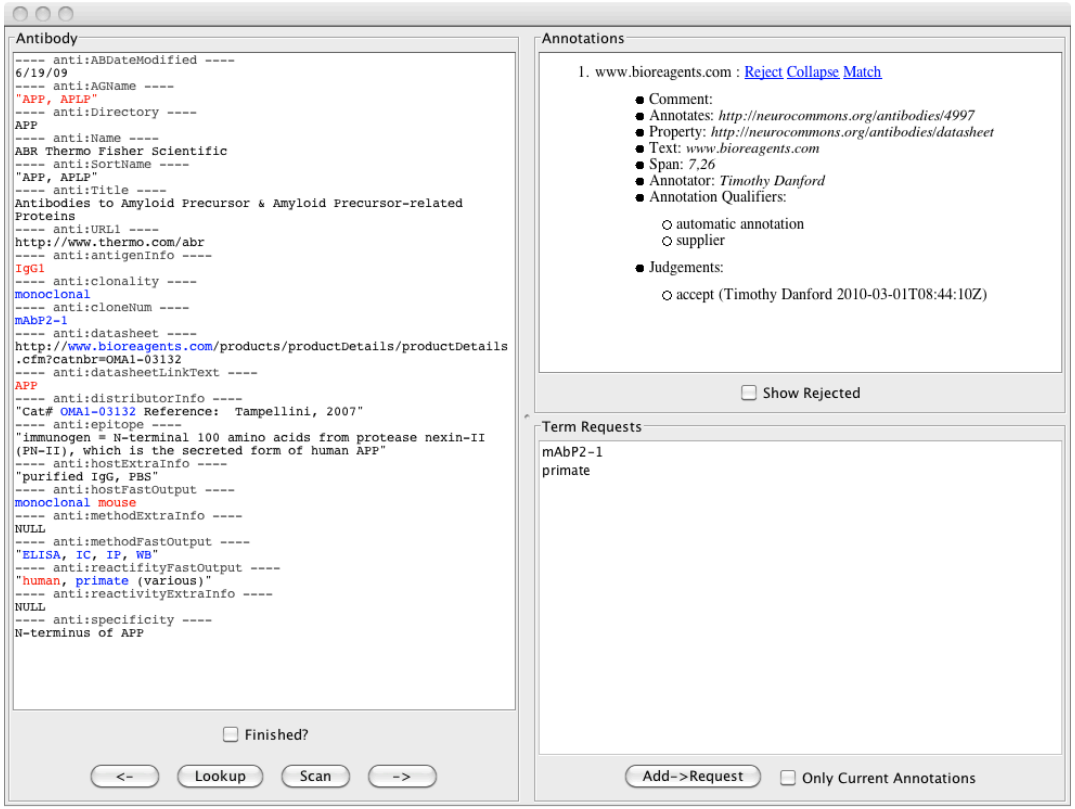
Antibodies modeling

- The 'term broker' can be used for improving our antibodies modeling process
- We are starting with 3k antibodies thaken from the AlzForum database (20k+ antibodies) (http://www.alzforum.org/res/com/ant/)
- The modeling starts with semi-automatic annotation of the record through our tool

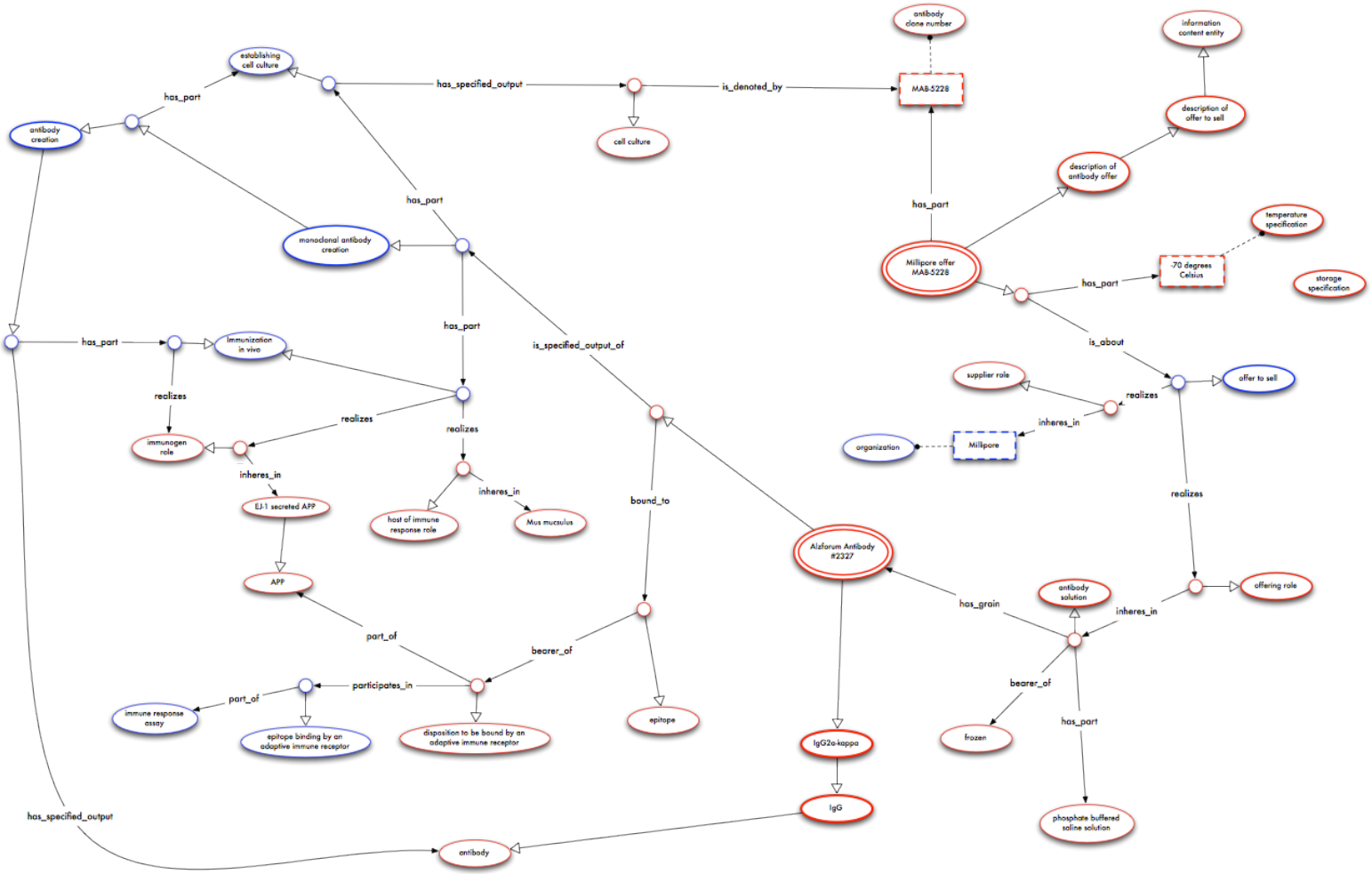
Antibodies to Amyloid Precursor & Amyloid Precursor-related Proteins						
From the Manufacturers						
Antigen (Clone) IgG	Distributor (Producer)	Immunogen (Epitope)	Host (Formulation)	Methods glossary	Specificity	Reactivity top
APP (3E9) IgG_1	Leinco Technologies cat# A121	immunogen = APP695 synthetic peptide (L E V P T D G N A G L L A E P Q I A M F C)	monoclonal mouse (lyophilized, PBS, sucrose, affinity purified)	IH(P), WB	APP695, specific sequence	human, mouse (other species not tested)

Annotation Extracts Structured Information from Antibody Records

- Automatic dictionaries identify regions of structured text
- Java application allows curators to review, reject, or extend annotations
- Annotations exported to an OBI RDF format
- Terms requests submitted as batch request



Annotations Translated Into OBI-Compliant RDF



Terms requests submitted as batch request

- We need to generate temporary IDs to be included in the request
- Temporary IDs are used in the application for regular functioning and for keeping track of the requests
- We need coded responses for the term requests in order to process them automatically in the application

People

Mass General Hospital

- Paolo Ciccarese
- Timothy Danford
- Tim Clark
- Sudeshna Das

Science Commons

- Alan Ruttenberg
- Jonathan Rees
- Kaitlin Thaney

Alzforum

- Elizabeth Wu
- Gwen Wong