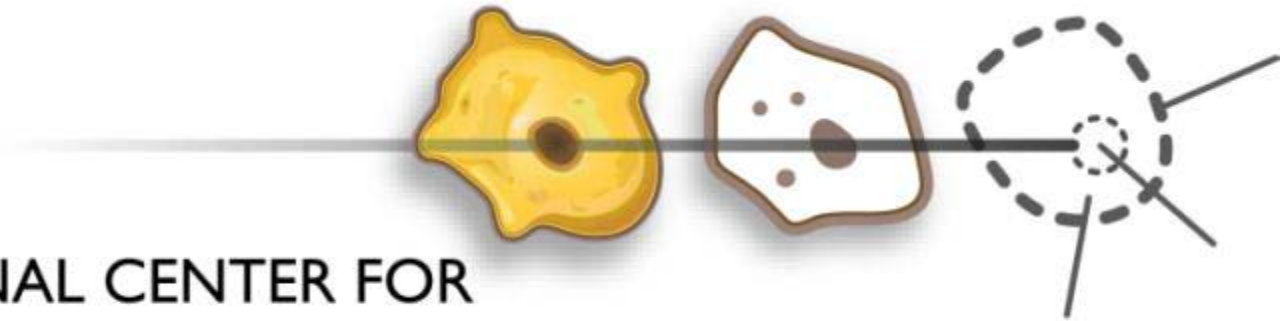
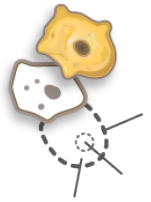
Ontology-oriented resources from the National Center for Biomedical Ontology

Mark Musen and Nigam Shah
{musen, nigam}@stanford.edu



NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

NCBO: Key activities



- We **create and maintain a library** of biomedical ontologies.
- We **build tools and Web services** to enable the use of ontologies and their derivatives.
- We **collaborate with scientific communities** that develop and use ontologies.

THE NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

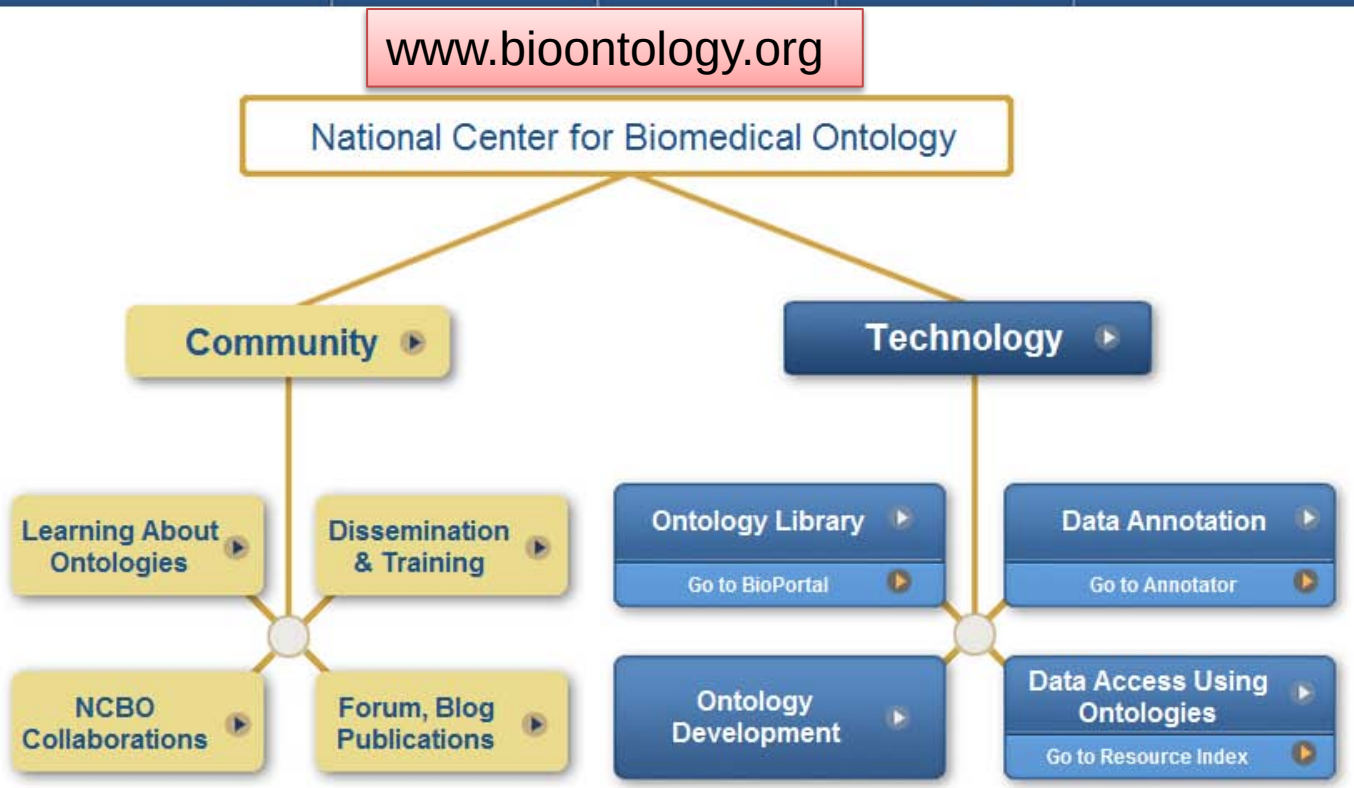
Search | Search Site | Search Google | LOGIN

Home | News & Events | Dissemination & Training | Collaboration | Community | Technology | About Us

Of Current Interest

- **Webinar**: Webinar series takes summer break and will resume on Wednesday, Sept. 1, 10 AM (PST): Cecilia Arighi, Protein Information Resource
- **Recent Publication**: Jonquet, C, Musen, MA, Shah, NH (2010): Building a biomedical ontology recommender web service
- **Recent Release**: BioPortal 2.5 (June 7, 2010)
- **Call for Panels**: Panels due October 18, 2010, AMIA 2011 Summit on Translational Bioinformatics
- **NCBO - Public Talks - Subscribe**
- More News & Events
- More Webinars
- **Follow us on Twitter**

www.bioontology.org



NCBO User Profile

Gary An
The University of Chicago

"The development of agent-based modeling is intimately tied to the advances in the ontologies and tools provided by NCBO via BioPortal." ***More >***

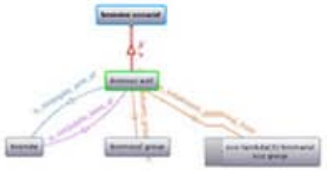
Other profiles >

Video

Learn about Biomedical Ontologies. Watch an introductory video.

Browse ontologies in BioPortal!

BioPortal allows users to browse, search and visualize ontologies.



Access all ontologies that are available in BioPortal: You can filter this list by category to display ontologies relevant for a certain domain (try selecting "Anatomy or Experimental Conditions" in the "Filter by category" menu below). You can also filter ontologies that belong to a certain group such as ontologies from the OBO Foundry, or ontologies from the Cancer Biomedical Informatics Grid (caBIG). Subscribe to the BioPortal RSS feed to receive alerts for submissions of new ontologies, new versions of ontologies, new notes, and new projects. You can subscribe to feeds for a specific ontology at the individual ontology page. Add a new biomedical ontology to BioPortal using the Submit New Ontology link (you need to sign in to see this link).

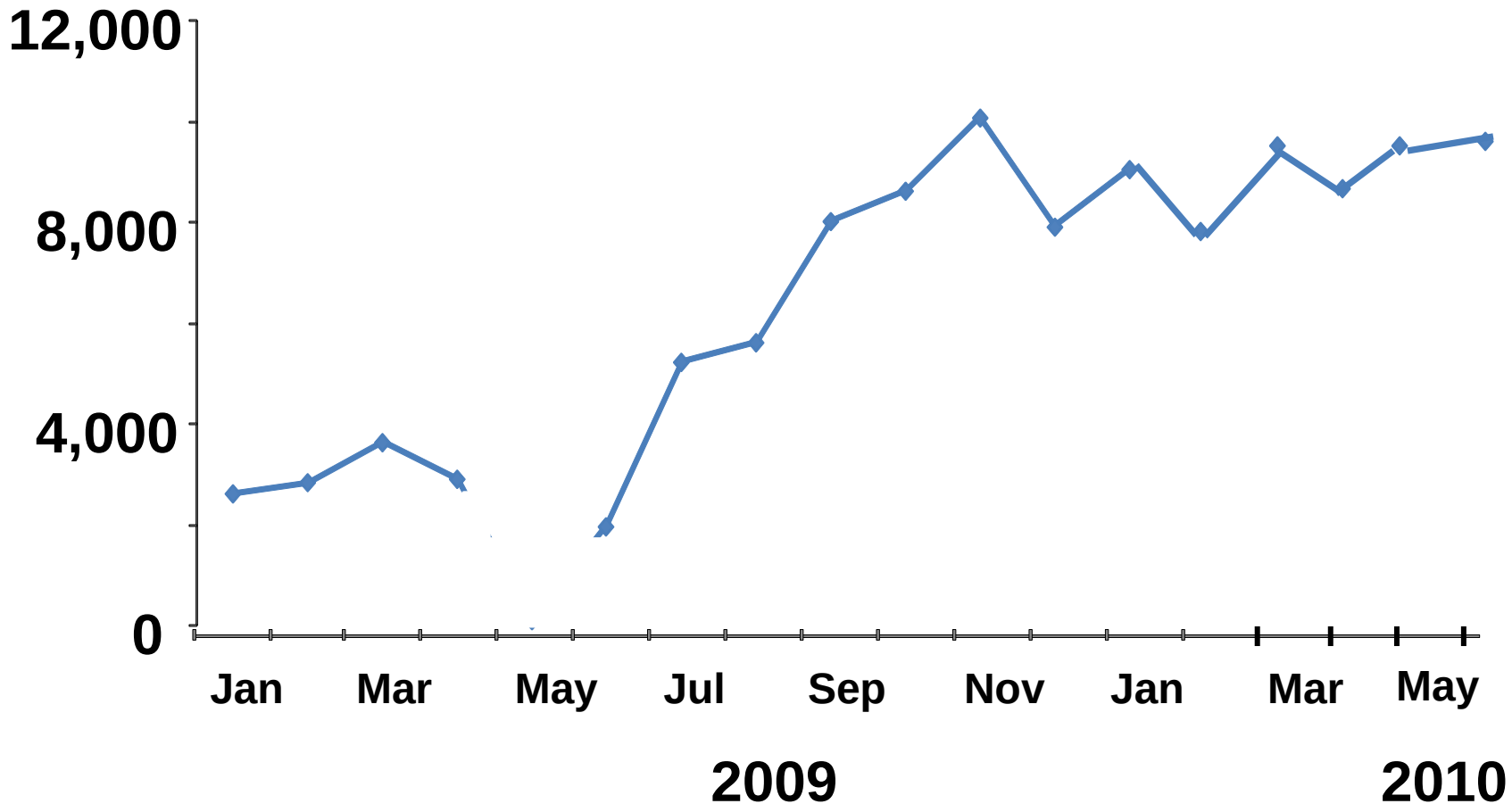
SUBMIT ONTOLOGY	Submit New Ontology
FILTER BY CATEGORY	All Categories
FILTER BY GROUP	All Groups — Link To This Filter
FILTER BY TEXT	

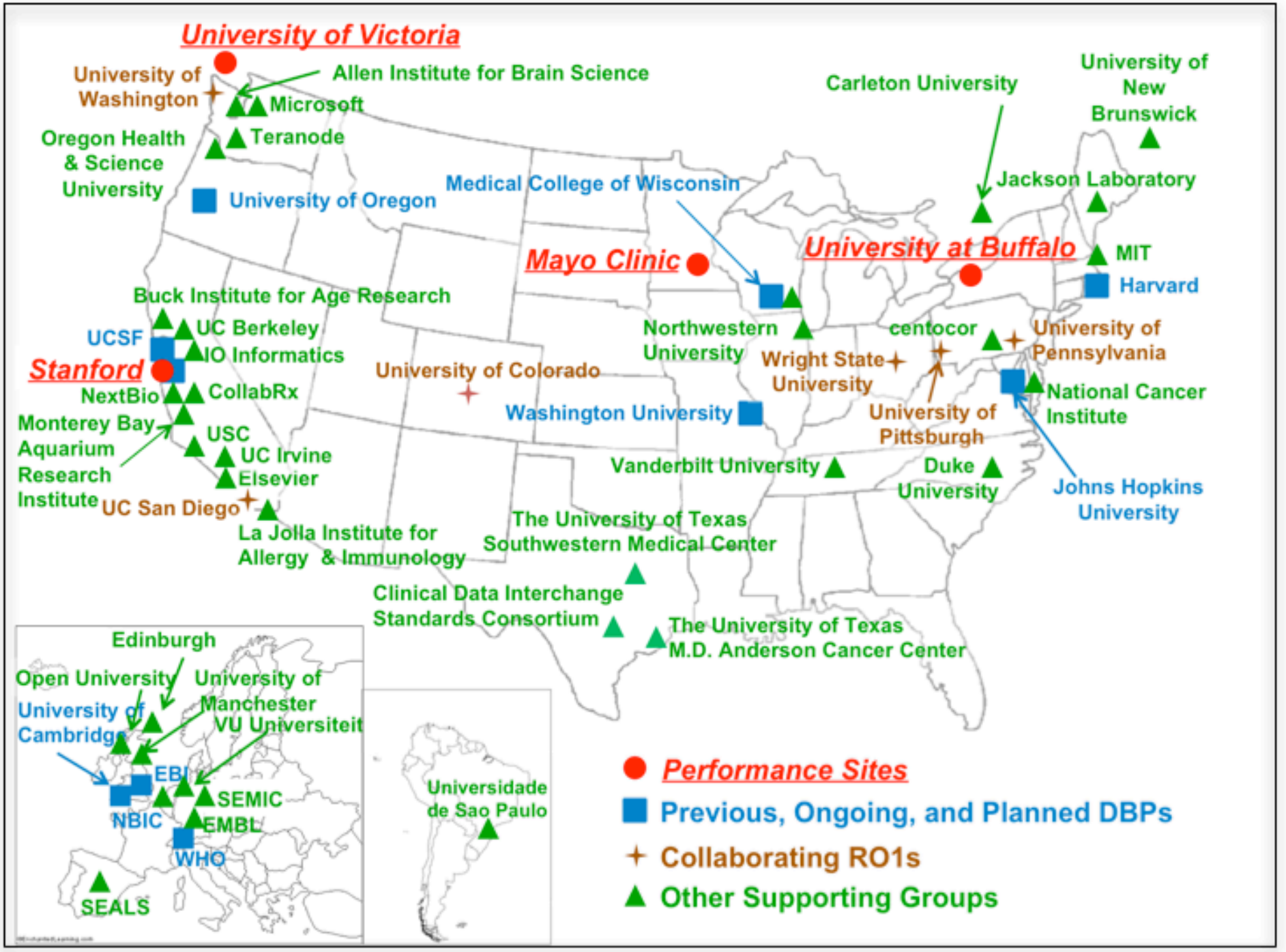
Subscribe to all updates

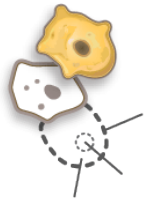
Go to BioPortal

ONTOLOGY NAME	FORMAT	VERSION	AUTHOR	UPLOADED ON	GROUP	STATUS
ABA Adult Mouse Brain (ABA)	OWL	1.0	Allen Institute for Brain Science	08/08/2009		Explore
African Traditional Medicine (ATMO)	OBO Format	1.101	Ghislain Atemezing	06/28/2009		Explore
Amino Acid (amino-acid)	OWL	1.2	Nick Drummond, Georgina Moulton, Robert Stevens, Phil Lord	04/25/2009		Explore
Amphibian gross anatomy (AAO)	OBO Format	1.8	AmphiAnat list	07/30/2008	OBO Foundry	Explore
Amphibian taxonomy (ATO)	OBO Format	See Remote Site	AmphiAnat list	08/14/2009		
Animal natural history and life history (ADW)	PROTEGE	See Remote Site	Http://animaldiversity Administrators	04/27/2009		
Ascomycete phenotype ontology (APO)	OBO Format	1.6	Fungal_anatomy Administrators	09/01/2009		Explore
Basic Formal Ontology (BFO)	OWL	1.1		07/24/2009		Explore
Basic Vertebrate Anatomy (basic-vertebrate-gross-anatomy)	OWL	1.1		01/16/2007		Explore
Bilateria anatomy (BILA)	OBO Format	See Remote Site	Thorsten Heinrich	04/13/2009		
Biological imaging methods (FBbi)	OBO Format	1.1	Image Administrators	07/30/2008		Explore

Total Monthly Visits to BioPortal




University of Victoria
University of Washington
Allen Institute for Brain Science
Microsoft
Teranode
Oregon Health & Science University
University of Oregon
Medical College of Wisconsin
Carleton University
University of New Brunswick
Jackson Laboratory
Mayo Clinic
University at Buffalo
MIT
Harvard
Buck Institute for Age Research
UC Berkeley
UCSF
IO Informatics
Stanford
NextBio
CollabRx
Monterey Bay Aquarium Research Institute
USC
UC Irvine
Elsevier
UC San Diego
La Jolla Institute for Allergy & Immunology
University of Colorado
Northwestern University
Wright State University
centocor
University of Pennsylvania
National Cancer Institute
Washington University
University of Pittsburgh
Vanderbilt University
Duke University
Johns Hopkins University
The University of Texas Southwestern Medical Center
Clinical Data Interchange Standards Consortium
The University of Texas M.D. Anderson Cancer Center
Edinburgh
Open University
University of Manchester
University of Cambridge
VU Universiteit
EBI
SEMIC
NBIC
EMBL
WHO
SEALS
Universidade de Sao Paulo
Performance Sites
Previous, Ongoing, and Planned DBPs
Collaborating RO1s
Other Supporting Groups



PART-I

http://rest.bioontology.org

Ontology Services

Views

- Download
- Traverse
- Search
- Comment

Mapping Services

- Create
- Download
- Upload

Widgets

- Tree-view
- Auto-complete
- Graph-view

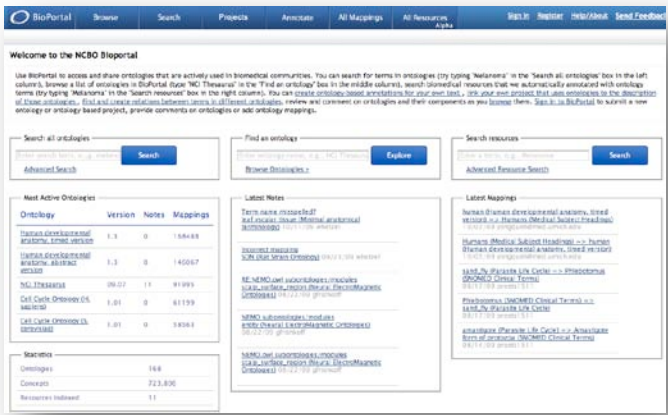
Annotation Services

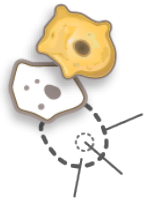
Concept recognition

Data Services

Fetch “data” annotated with a given concept

http://bioportal.bioontology.org





ONTOLOGY SERVICES

Accessing, browsing, searching and traversing ontologies in *Your* application

www.bioontology.org/wiki/index.php/NCBO_REST_services

Contents [hide]

1 Announce List
2 Services to access ontologies and ontology versions
- 2.1 List all the latest version of ontologies
- 2.2 Get a specific ontology based on a version id
- 2.3 Download an ontology file
- 2.4 Download the latest ontology file
- 2.5 Get all versions of an ontology from a virtual ontology id
- 2.6 Get latest version of an ontology id
- 2.7 List all ontology categories
- 2.8 List all ontology groups

3 Services to access ontology views and ontology view versions
- 3.1 List all the latest version of views
- 3.2 Get all versions of views from a virtual ontology id

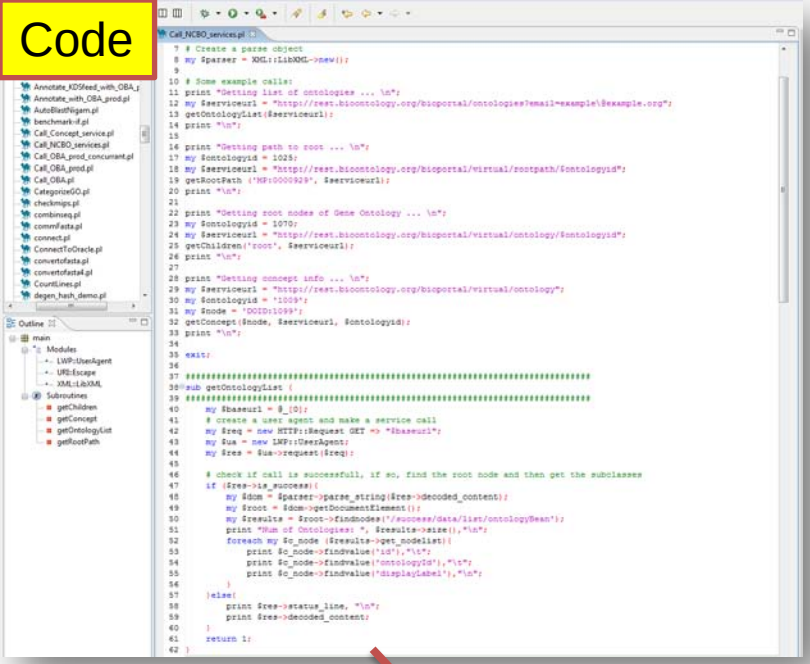
4 Concept services
- 4.1 Get concept
- 4.2 Get all root concepts for an ontology id
- 4.3 Get concept for latest ontology version id
- 4.4 Get all concepts

5 Search services
- 5.1 Search BioPortal

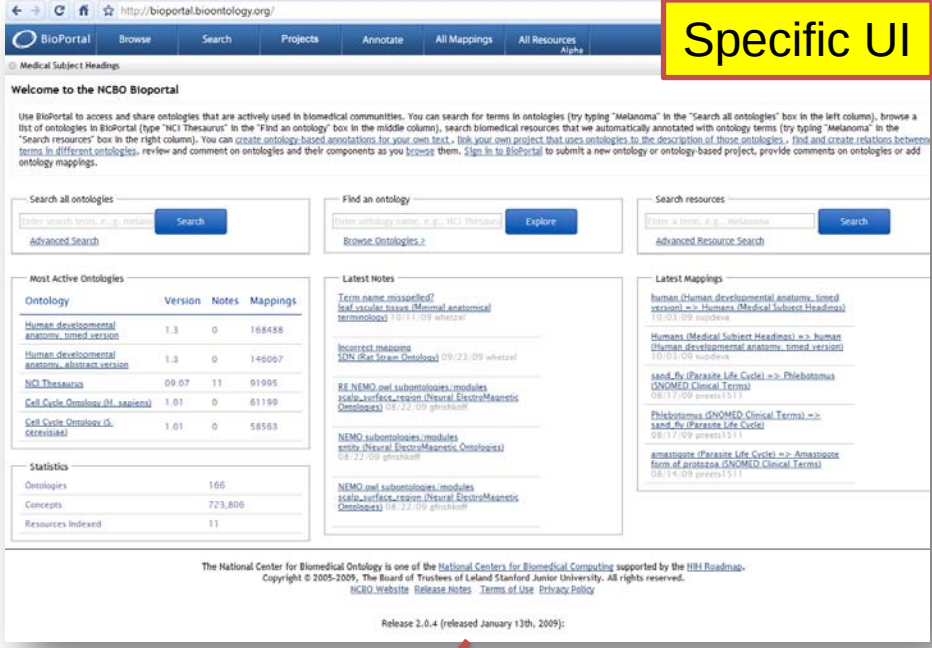
6 Hierarchy Services
- 6.1 Get parents/children of a given concept in a specific ontology version
- 6.2 Get parents/children of a given concept in the latest version of a given ontology
- 6.3 Get paths to roots/leaves from a concept in a specific ontology version
- 6.4 Get paths to root/leaves from a concept in the latest version of a given ontology
- 6.5 Get siblings of a given concept in a specific ontology version
- 6.6 Get siblings of a given concept in the latest version of a given ontology

7 Notes Service
8 Mapping Service
9 Annotator Service
10 Usage Logging
11 Overview - Using NCBO Technology in Your Project

Code



Specific UI



http://rest.bioontology.org/<SERVICE>

http://rest.bioontology.org/bioportal/ontologies

```
<success>
  <data>
    <list>
      <ontologyBean>
        <id>40401</id>
        <ontologyId>1351</ontologyId>
        <displayLabel>Medical Subject Headings</displayLabel>
        <description>Medical Subject Headings (MeSH);National L
        <abbreviation>MSH</abbreviation>
        <format>RRF</format>
        <internalVersionNumber>1</internalVersionNumber>
        <versionNumber>2009_2009_02_13</versionNumber>
        <contactName>Stuart Nelson, M.D.</contactName>
        <contactEmail>nelson@nlm.nih.gov</contactEmail>
        <statusId>3</statusId>
        <categoryIds>
          <int>5058</int>
        </categoryIds>
        <groupIds>
          <int>6008</int>
        </groupIds>
        <isFoundry>0</isFoundry>
        <dateCreated>2009-07-31 06:01:06.0 PDT</dateCreated>
```

Bethesda, MD;ENG</description>

description>

```
        <versionNumber>1.194</versionNumber>
        <versionStatus>production</versionStatus>
        <contactName>Anatomy JAX</contactName>
        <contactEmail>anatomy@informatics.jax.org</contactEmail>
        <statusId>3</statusId>
        <categoryIds>
          <int>2812</int>
          <int>2811</int>
          <int>2810</int>
          <int>2817</int>
        </categoryIds>
```

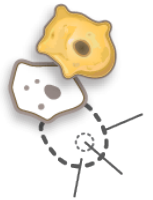
http://rest.bioontology.org/bioportal/search/melanoma/?ontologyids=1351

```
<success>
  <accessedResource>/bioportal/search/melanoma/</accessedResource>
  <accessDate>2009-10-14 15:12:32.410 PDT</accessDate>
  <data>
    <page>
      <pageNum>1</pageNum>
      <numPages>1</numPages>
      <pageSize>85</pageSize>
      <numResultsPage>85</numResultsPage>
      <searchBean>
        <ontologyVersionId>40401</ontologyVersionId>
        <ontologyId>1351</ontologyId>
        <ontologyDisplayLabel>Medical Subject Headings</ontologyDisplayLabel>
        <recordType>RECORD_TYPE_PREFERRED_NAME</recordType>
        <conceptId>D008545</conceptId>
        <conceptIdShort>D008545</conceptIdShort>
        <preferredName>Melanoma</preferredName>
        <contents>Melanoma</contents>
      </searchBean>
      <searchBean>
        <ontologyDisplayLabel>Medical Subject Headings</ontologyDisplayLabel>
        <recordType>RECORD_TYPE_PREFERRED_NAME</recordType>
        <conceptId>D008546</conceptId>
        <conceptIdShort>D008546</conceptIdShort>
        <preferredName>Melanoma, Experimental</preferredName>
        <contents>Melanoma, Experimental</contents>
      </searchBean>
      <searchBean>...</searchBean>
      <searchBean>...</searchBean>
      <searchBean>...</searchBean>
      <searchBean>...</searchBean>
      <searchBean>...</searchBean>
      <searchBean>
        <ontologyVersionId>40401</ontologyVersionId>
        <ontologyId>1351</ontologyId>
        <ontologyDisplayLabel>Medical Subject Headings</ontologyDisplayLabel>
        <recordType>RECORD_TYPE_SYNONYM</recordType>
        <conceptId>C092706</conceptId>
        <conceptIdShort>C092706</conceptIdShort>
        <preferredName>BAGE protein, human</preferredName>
        <contents>B melanoma antigen, human</contents>
      </searchBean>
```

http://rest.bioontology.org/bioportal/virtual/ontology/1351/D008545

```
<success>
  <accessedResource>/bioportal/virtual/ontology/1351/D008545</accessedResource>
  <accessDate>2009-10-14 15:05:21.635 PDT</accessDate>
  <data>
    <classBean>
      <id>D008545</id>
      <label>Melanoma</label>
      <relations>
        <entry>
          <string>UMLS_CUI</string>
          <list>
            <string>C0025202</string>
          </list>
        </entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>
          <string>SYNONYM PM</string>
          <list>
            <string>Melanomas</string>
            <string>Malignant Melanomas</string>
            <string>Melanomas, Malignant</string>
            <string>Melanoma, Malignant</string>
          </list>
        </entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>...</entry>
        <entry>
          <string>Definition</string>
          <list>
            <string>A malignant neoplasm derived from cells that are capable of forming melanin, which may occur in the skin of any part of the body
            alia, anus, oral cavity, or other sites. It occurs mostly in adults and may originate de novo or from a pigmented nevus or malignant len
            l lymph nodes, liver, lungs, and brain are likely to be involved. The incidence of malignant skin melanomas is rising rapidly in all par
             of Dermatology, 4th ed, p2445)</string>
          </list>
```

Wikipathways uses Ontology Services



BETA
WIKIPATHWAYS
Pathways for the People

search

Google™ Custom Searc

navigation
- Home
- Help

pathway
- Create
- Browse
- Wish List
- Download

overview
- Recent Changes
- Most Viewed
- Most Edited
- New Pathways

community
- About us
- Contact us
- How to cite
- GenMAPP Portal
- BiGCaT Portal
- Micronutrient Portal
- Development

toolbox

special

Ontology Index

List | Image | Tree

Species :
- All Species
- Anopheles gambiae
- Arabidopsis thaliana
- Bos taurus
- Bacillus subtilis
- Caenorhabditis elegans
- Canis familiaris
- Danio rerio
- Drosophila melanogaster
- Escherichia coli
- Equus caballus
- Gallus gallus
- **Homo sapiens**
- Mus musculus
- Oryza sativa
- Pan troglodytes
- Rattus norvegicus
- Saccharomyces cerevisiae
- Xenopus tropicalis

Ontologies :
Pathway Ontology
- **regulatory pathway (42)**
- signaling pathway (4)
- classic metabolic pathway (19)
- disease pathway (1)

Disease
- temp holding (2)

- regulatory pathway (42)
  - homeostasis pathway (0/6)
  - cell death pathway (0/6)
  - pathways of replication, repair, gene expression, protein biosynthesis (0/15)
    - cell cycle checkpoint pathway (1/1)
    - translation pathway (1/1)
    - transcription pathway (5/5)
    - DNA repair pathway (3/5)
      - >> ***Homologous recombination***
      - >> ***Mismatch repair***
      - >> ***Non-homologous end joining***
  - transport pathway (8/8)
  - immune response pathway (4/7)

Ontology Tags

Pathway Ontology : cell adhesion signaling pathway, immune response pathway, inflammatory response pathway
Disease : Signal Transduction Pathway Deregulation, disease of response to stimulus
Cell Type : lymph gland plasmatocyte

dna repair

- **altered DNA repair pathway** — Pathway ontology
- **DNA repair pathway** — Pathway ontology
- **Abnormal DNA Repair** — Human disease
- **DNA Repair Deficiency** — Human disease

To add a tag, either select from the available ontology trees below or type a search term in the search box.

Disease Cell Type

Ontology Tags

Pathway Ontology : cell adhesion signaling pathway, immune response pathway, inflammatory response pathway
Disease : Signal Transduction Pathway Deregulation, disease of response to stimulus
Cell Type : lymph gland plasmatocyte

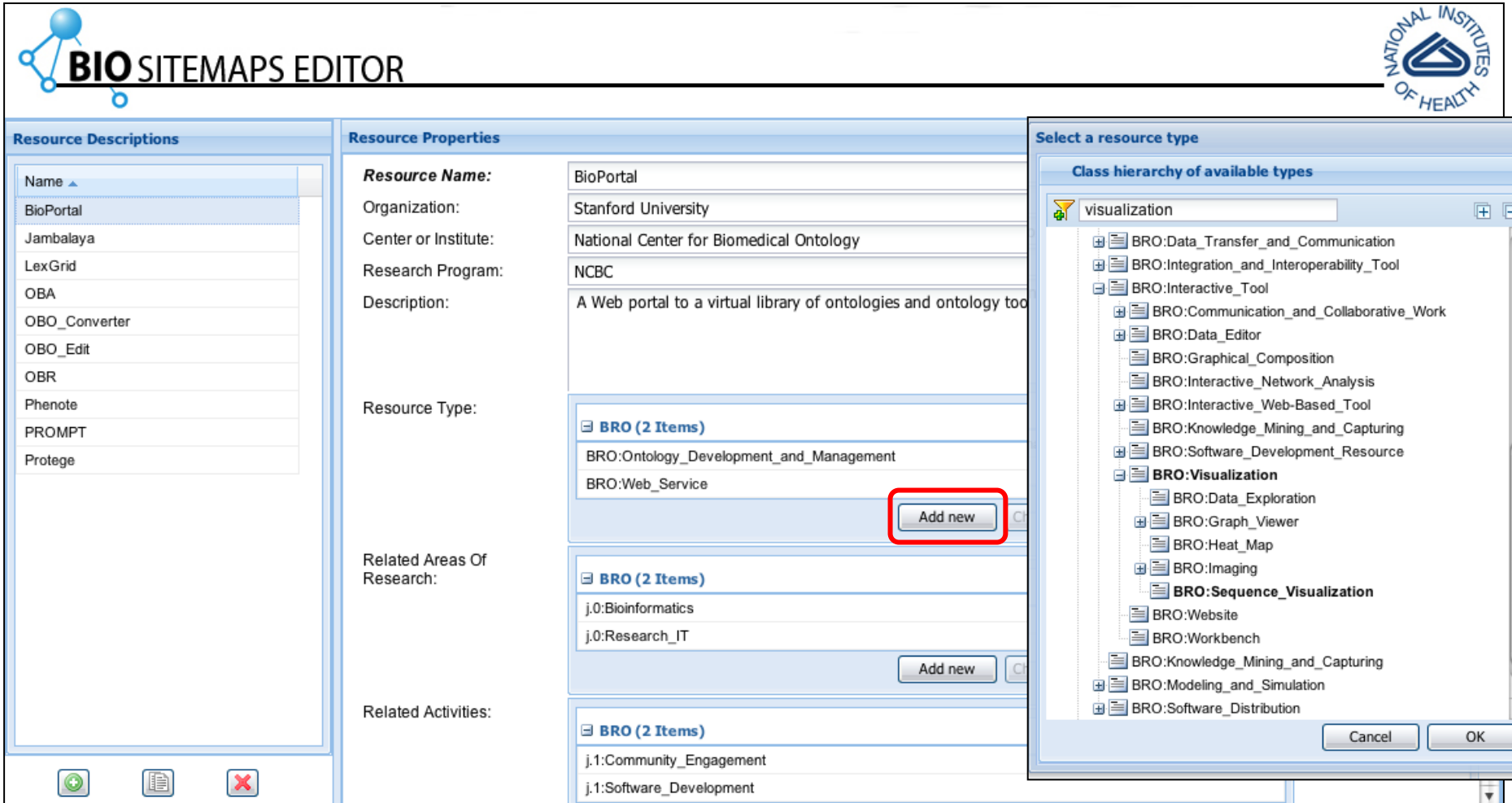
Term : altered DNA repair pathway
ID : PW:0000292

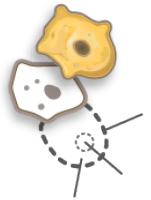
altered DNA repair pathway

To add a tag, either select from the available ontology trees below or type a search term in the search box.

Pathway Ontology Disease Cell Type

Biositemaps Editor

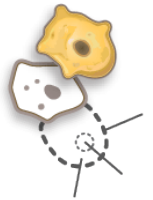




VIEWS

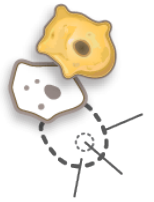
Custom subsets of large ontologies

Views and Value Sets



- Users can contribute their derivatives of BioPortal ontologies,
  - which become first-class objects in BioPortal and can be used as all other ontologies are (e.g., as value sets)
- Recently added: a view-extractor service
  - Enables users to extract a subtree of an ontology in OWL

Views in BioPortal



BioPortal | Browse | Search | Projects | Annotate | All Mappings | All Resources Alpha | Logged In As natasha My Account Log Out

SNOMED Clinical Terms

Subscribe to updates

Submit New Version | Edit Ontology Information

Metadata | Views (6) | Projects (1) | Reviews (0) | Metrics | Ontology Widgets

Create New View

Expand All | Collapse All

CORE Subset of SNOMED CT (Edit Metadata)

- **Description:** The CORE Problem List Subset of SNOMED CT
- **Ontology ID:** 2040
- **Definition Language:** Manual

VERSION	BASE VERSION	CREATED	CREATED BY	ONTOLOGY FILE	VISUALIZE
CORE_2009_01_31	2009_07_31	01/31/2009	Vivian A. Auld, auld@nlm.nih.gov	Download View	Explore

Submit New View Version

SNOMED Anatomy (Edit Metadata)

SNOMED Clinical Findings (Edit Metadata)

SNOMED Organism (Edit Metadata)

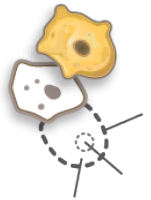
SNOMED Terminos Clinicos (Edit Metadata)

- **Description:** SNOMED Terminos Clinicos (SNOMED CT), Edicion en Espanol, Distribucion Internacional, Octubre de 2008, 2009_04_30
- **Ontology ID:** 1416

VERSION	BASE VERSION	CREATED	CREATED BY	ONTOLOGY FILE	VISUALIZE
2009_04_30	2009_07_31	04/30/2009	custserv@nlm.nih.gov	Download View	Explore
2009_04_30	2009_07_31	04/30/2009	custserv@nlm.nih.gov	Download View	Explore

Submit New View Version

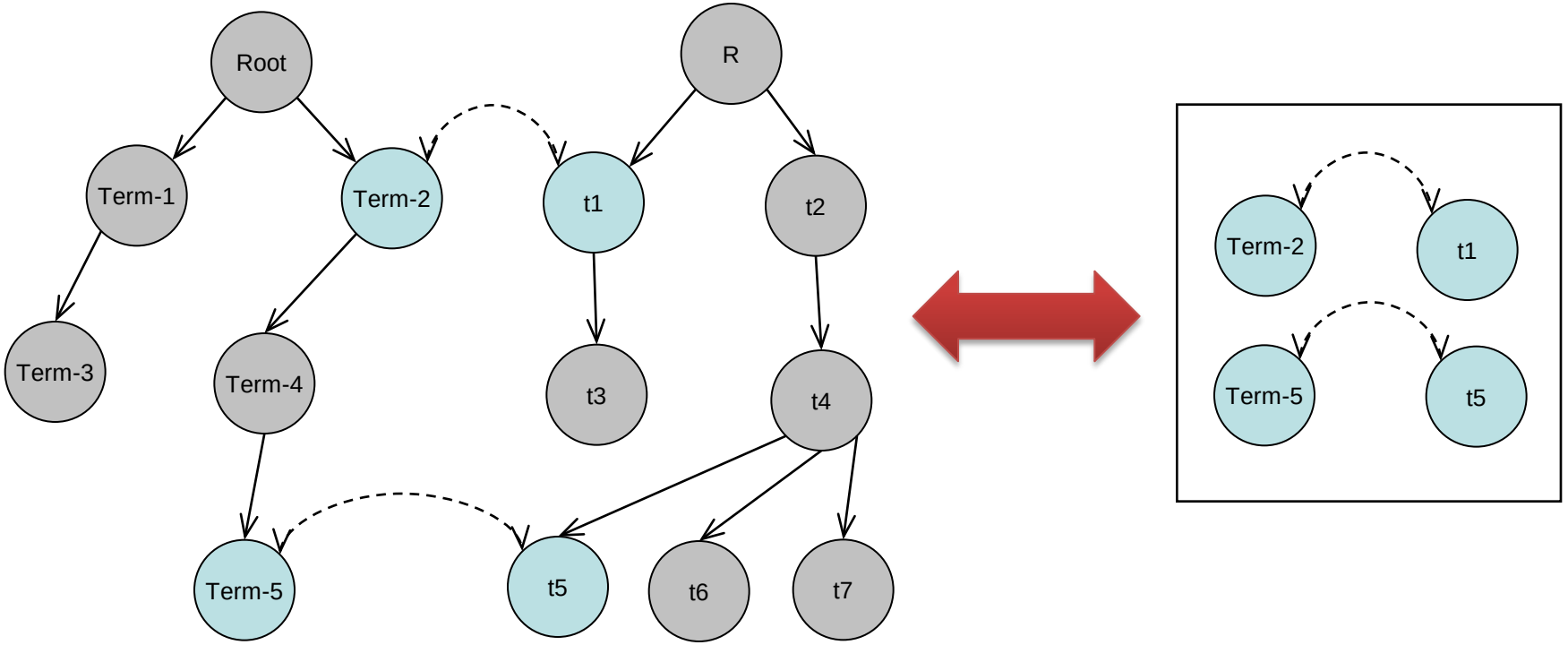
SNOMED Test Findings (Edit Metadata)



MAPPINGS

Using NCBO technology to integrate terminologies and ontologies

Mappings

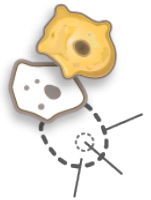


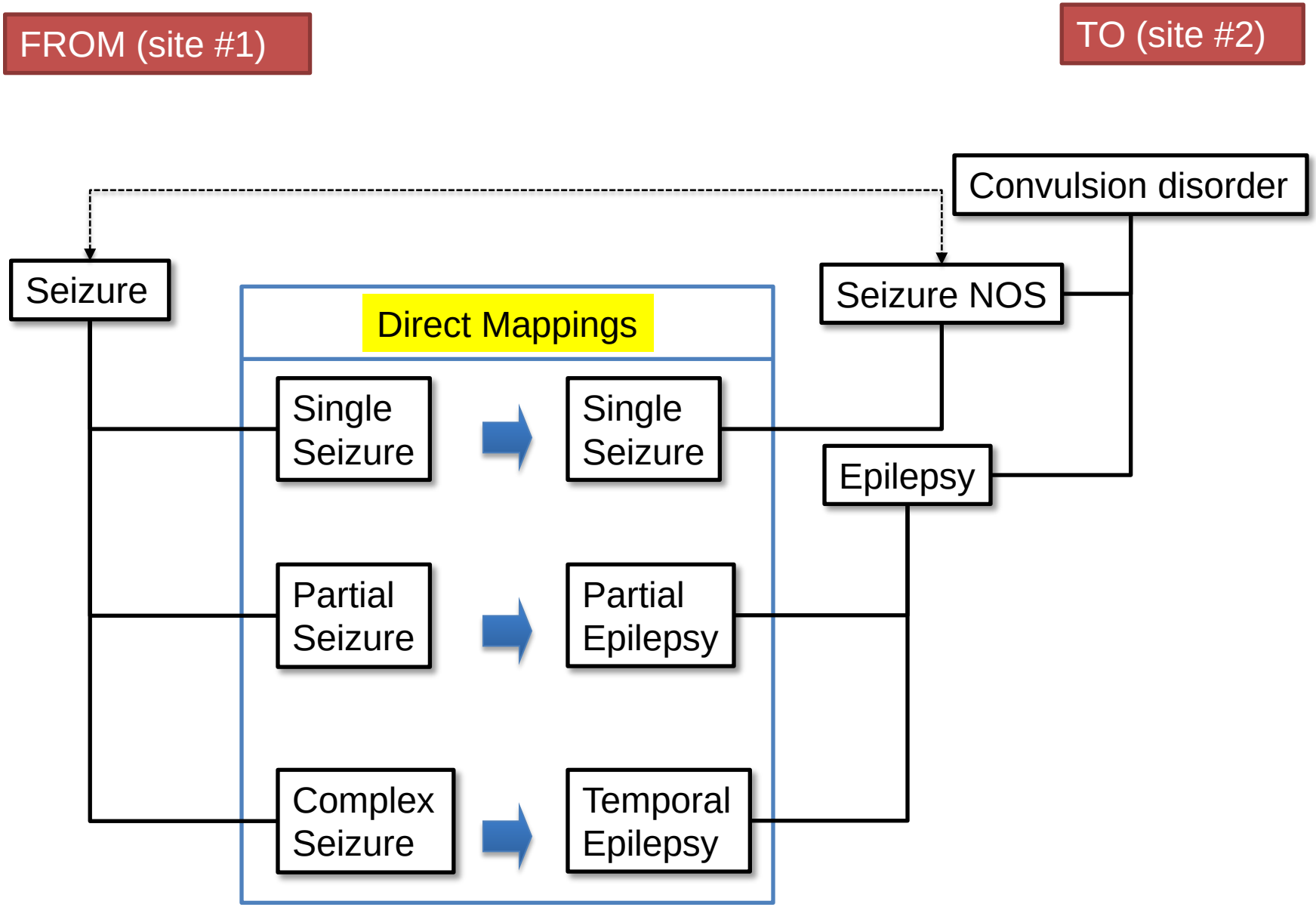
Ontology **A**

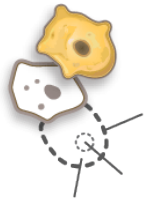
Ontology **B**

Upload or Download
mapping subsets

Using Mappings for query federation







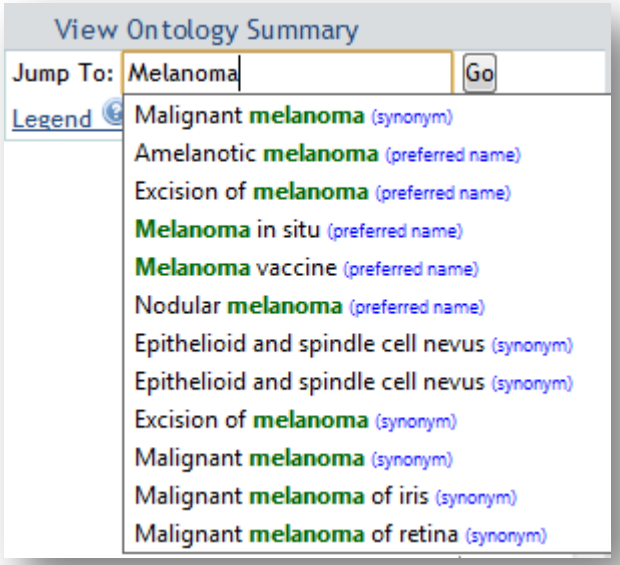
WIDGETS

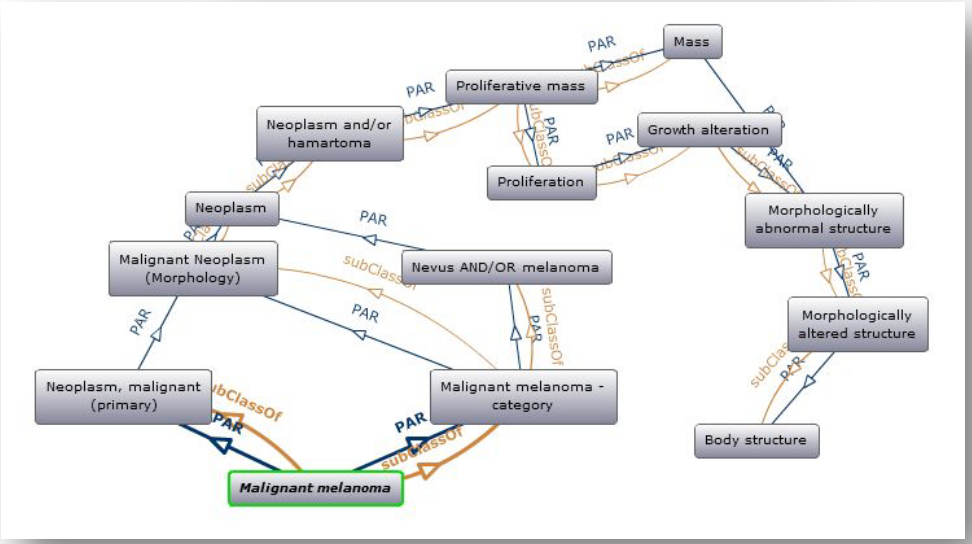
Using NCBO technology on your web pages

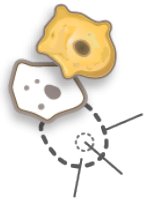
Ontology Widgets

UI components with "BioPortal inside":

- term-selection widget for a specific ontology
- form fields with auto-complete from a specific BioPortal ontology
- RSS feed for an ontology
- Visualization widget
- Tree widget





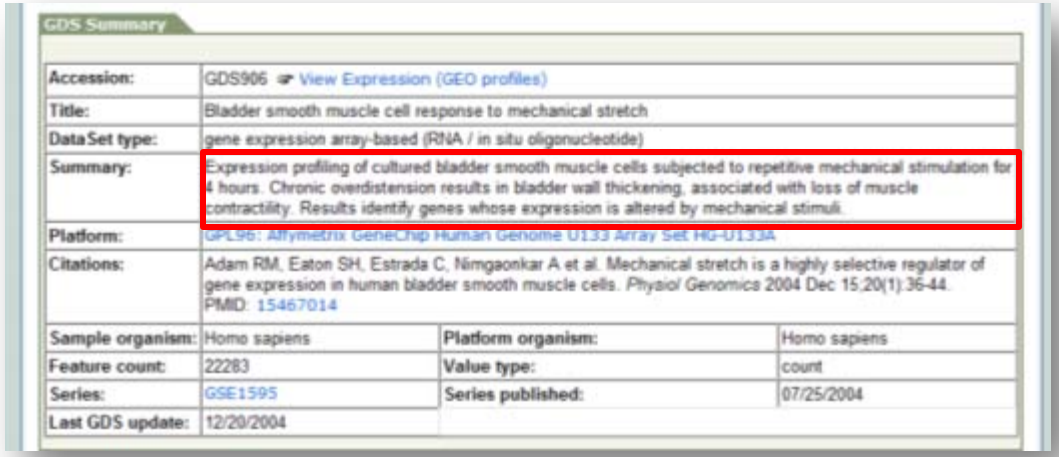


ANNOTATOR SERVICE

Using Ontologies to Annotate *Your* Data

Annotation as a Web service

Process textual metadata to automatically tag text with as many ontology terms as possible.

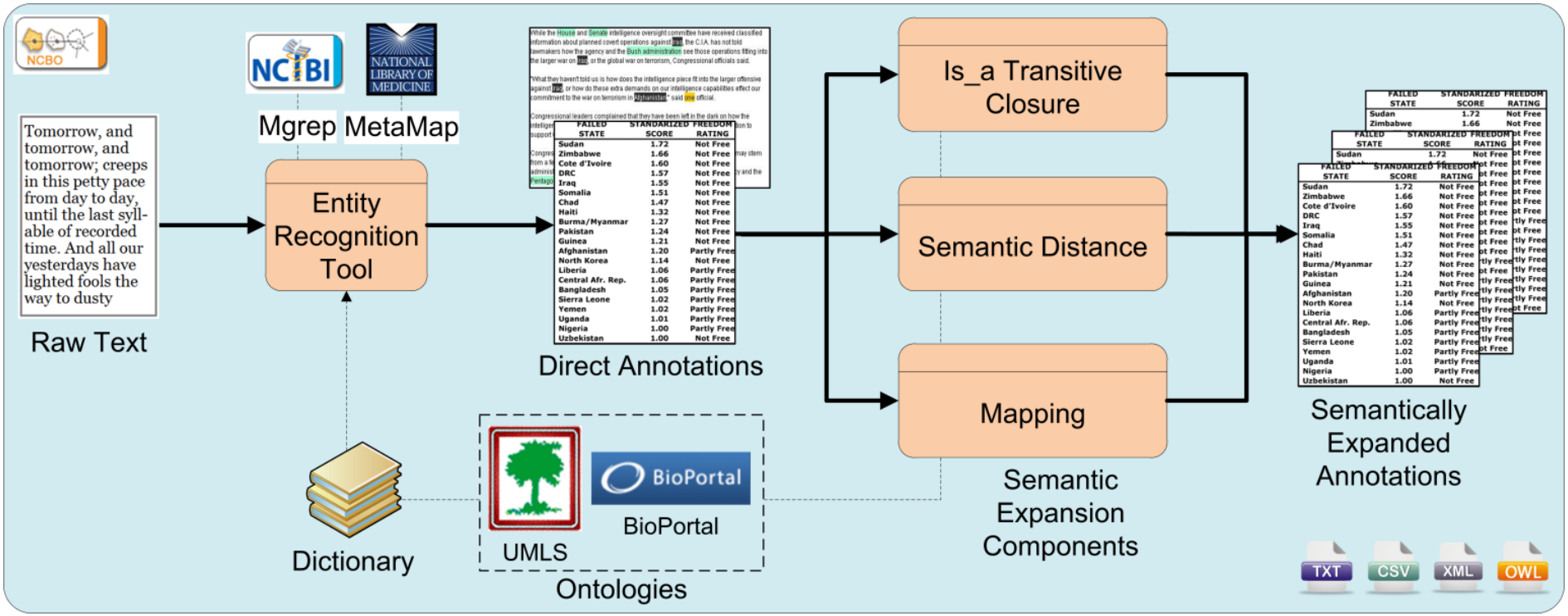


GDS Summary

Accession:	GDS906 View Expression (GEO profiles)		
Title:	Bladder smooth muscle cell response to mechanical stretch		
DataSet type:	gene expression array-based (RNA / in situ oligonucleotide)		
Summary:	Expression profiling of cultured bladder smooth muscle cells subjected to repetitive mechanical stimulation for 4 hours. Chronic overdistension results in bladder wall thickening, associated with loss of muscle contractility. Results identify genes whose expression is altered by mechanical stimuli.		
Platform:	GPL96: Affymetrix GeneChip Human Genome U133 Array Set HG-U133A		
Citations:	Adam RM, Eaton SH, Estrada C, Nimgaonkar A et al. Mechanical stretch is a highly selective regulator of gene expression in human bladder smooth muscle cells. *Physiol Genomics* 2004 Dec 15;20(1):36-44. PMID: 15467014		
Sample organism:	Homo sapiens	**Platform organism:**	Homo sapiens
Feature count:	22283	**Value type:**	count
Series:	GSE1595	**Series published:**	07/25/2004
Last GDS update:	12/20/2004		

Expression, Expression of bladder, bladder, smooth, bladder muscle, muscle, smooth muscle, cells, mechanical, mechanical stimulation, stimulation, Chronic, results, bladder overdistension, associated, associated with, with, loss, genes, altered

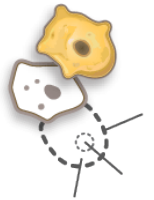
Annotator: workflow

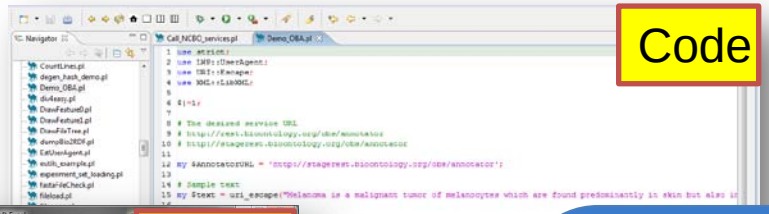


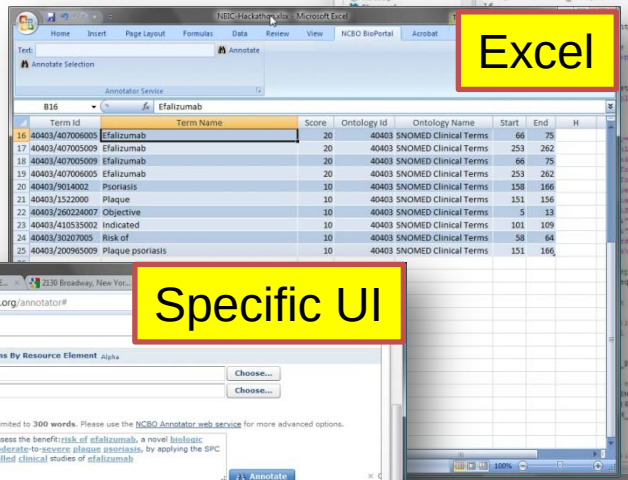
- "Melanoma is a malignant tumor of melanocytes which are found predominantly in skin but also in the bowel and the eye".
  - `39228/DOID:1909`, *Melanoma* in Human Disease

- Transitive closure
  - `39228/DOID:191`, *Melanocytic neoplasm*, direct parent of *Melanoma* in Human Disease
  - `39228/DOID:0000818`, *cell proliferation disease*, grand parent of *Melanoma* in Human Disease

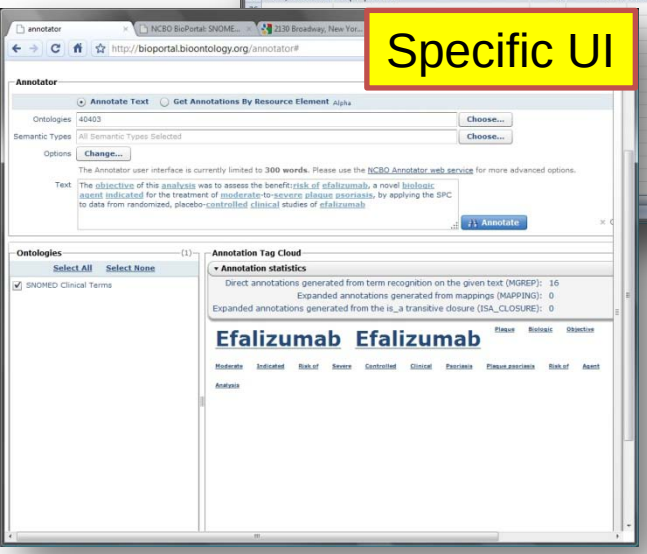
Multiple ways to access

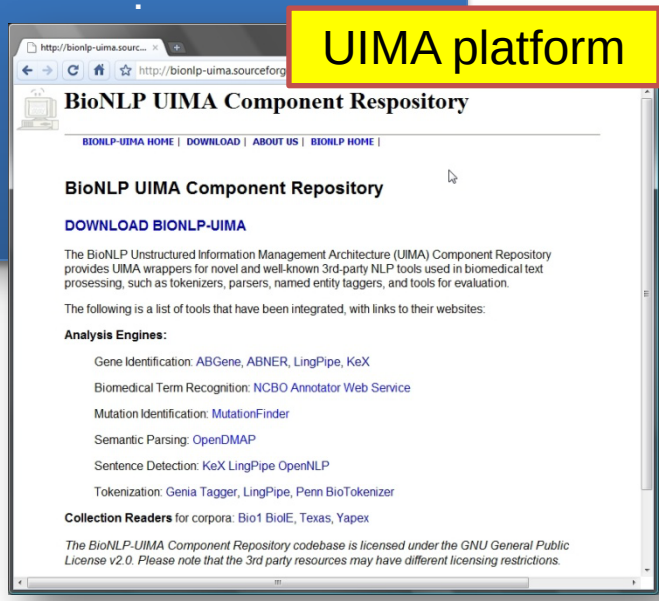


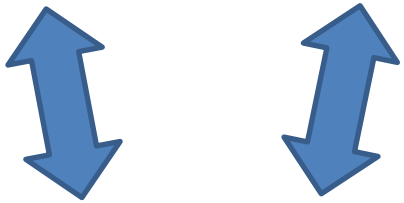




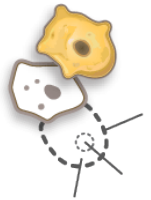
Word Add-in to call the Annotator Service
?







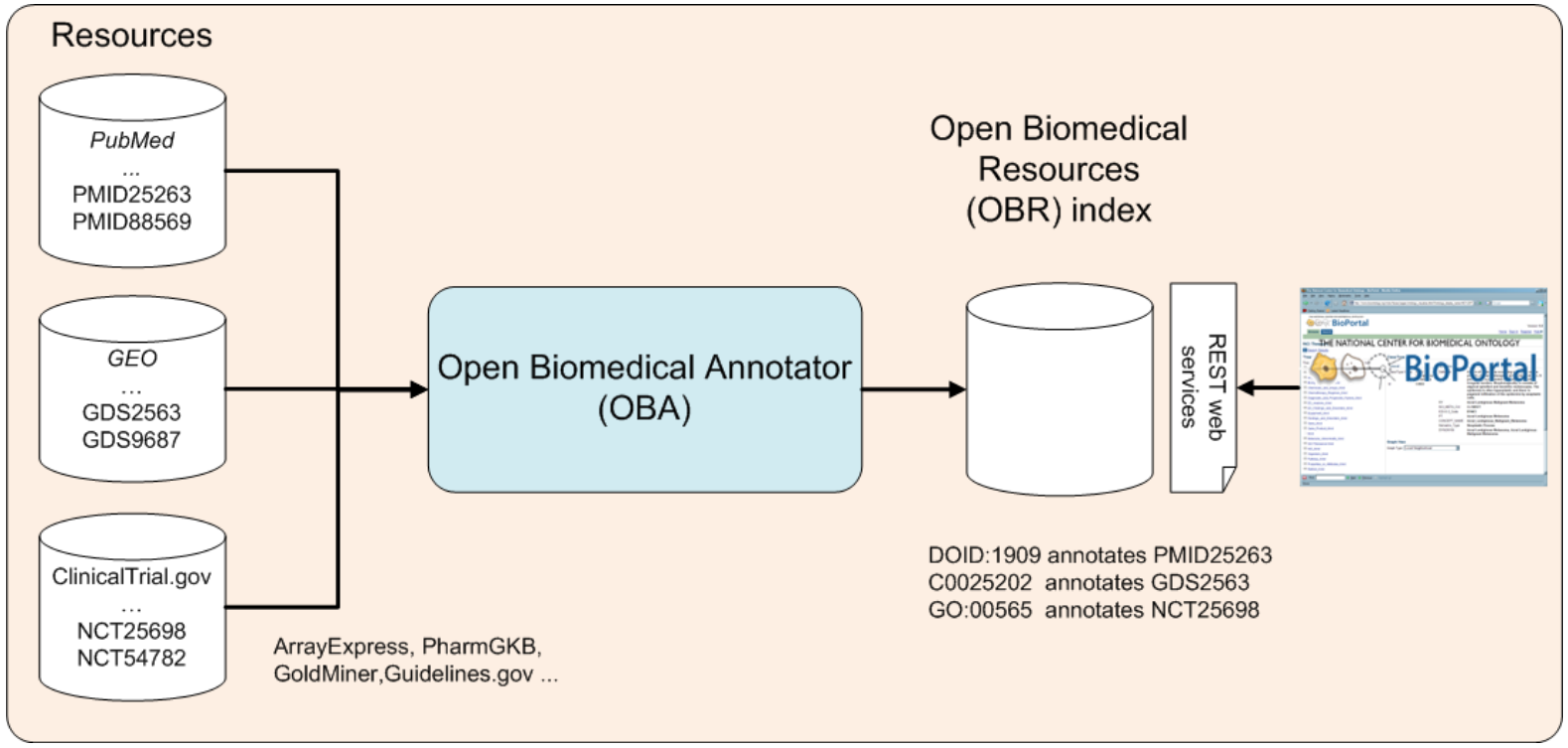
Annotator service



DATA SERVICE

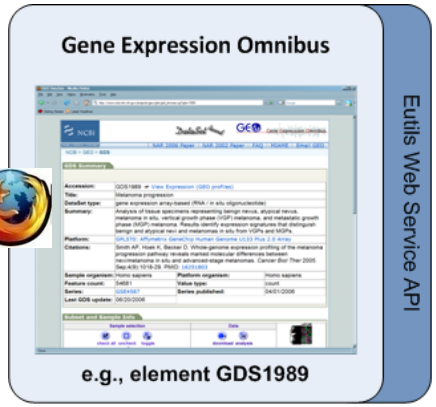
Using Ontologies to Access Public Data

Resource index: The Basic Idea



- The index can be used for:
  - Search
  - Data mining

Resources index: Example



Accessing resource elements

<title>
Melanoma progression
</title>

<description>
Analysis of tissue specimens representing benign nevus, atypical nevus, melanoma in situ, vertical growth phase (VGP) melanoma, and metastatic growth phase (MGP) melanoma. Results identify expression signatures that distinguish benign and atypical nevi and melanomas in situ from VGPs and MGPs.
</description>

Concept recognition

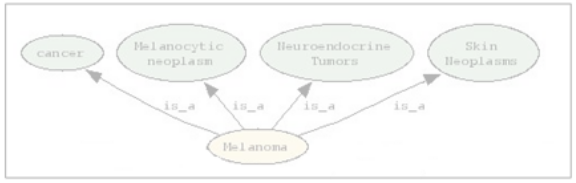
<title>
Melanoma progression
</title>

<description>
Analysis of tissue specimens representing benign nevus, atypical nevus, melanoma in situ, vertical growth phase (VGP) melanoma, and metastatic growth phase (MGP) melanoma. Results identify expression signatures that distinguish benign and atypical nevi and melanomas in situ from VGPs and MGPs.
</description>

23 direct annotations
(4 title, 19 description)

Example:
Melanoma, concept (DOID:1909) in ontology *Human disease*.

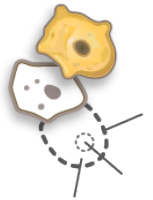
Semantic expansion



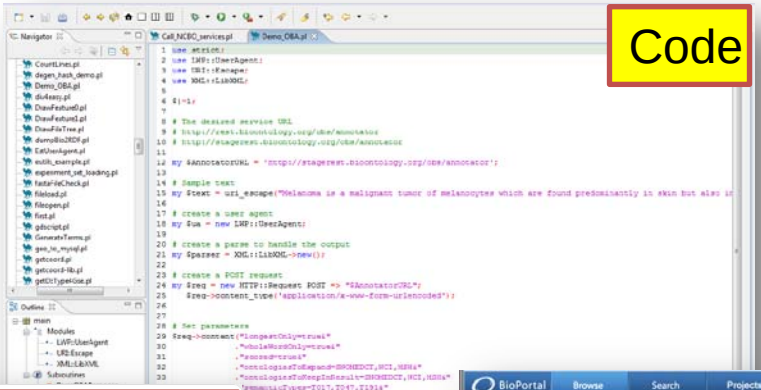
172 closure annotations

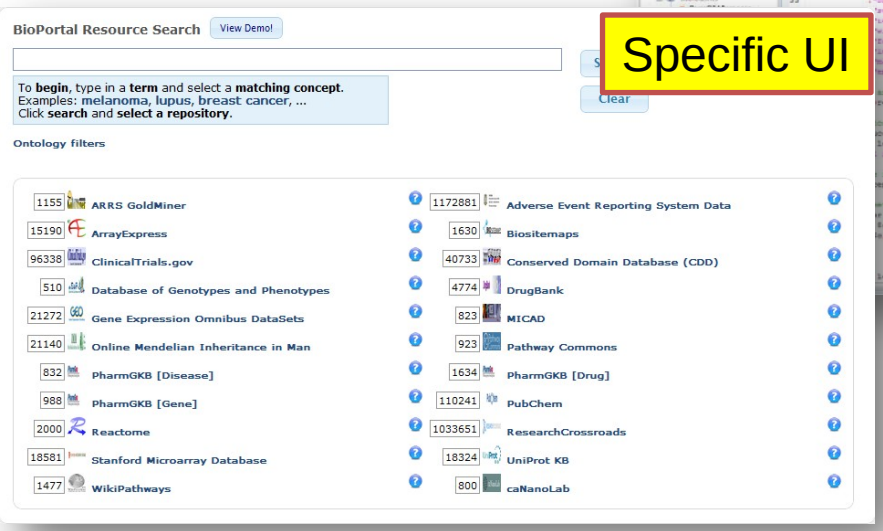
Examples:
Cancer, concept (DOID:162) in ontology *Human disease*
Skin Neoplasms, concept (DOID:3165) in ontology *Human disease*

Multiple ways to access

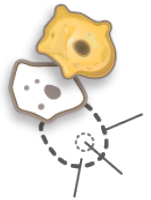


- Resources annotated = 22
- Total records = 3.5 million
- Direct annotations = … million
- After transitive closure = 16.4 Billion





Resource Index



PART-II

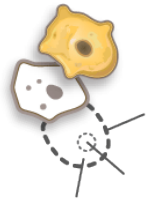
Use-cases based on ontology services

Sample user needs

- I need to restrict user input to a certain value set
- I need to extract the disease branch from SNOMEDCT
- I need to identify all terms mapped to UMLS CUI C0151779
- I need to code/annotate free-text with ontology terms
  - For data exchange, export to standard formats

Use-cases for users of i2b2

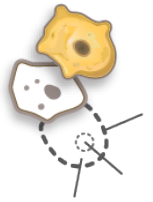
Aim 1: Integrate NCBO services in i2b2



- Preliminary results:
  - Export any ontology stored in BioPortal into the format used by i2b2's ontology cell

- Future Work:
  - Make the export code available as a service
  - Embed the extraction code into the i2b2 Ontology Cell to "pull" content
  - Ensure we have the latest versions of ontologies used by i2b2 and CTSA users (ICD9, ICD10, SNOMEDCT, RXNORM, LOINC, CPT)

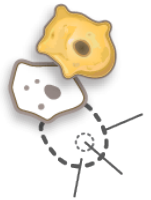
Aim 2: Mappings for query federation

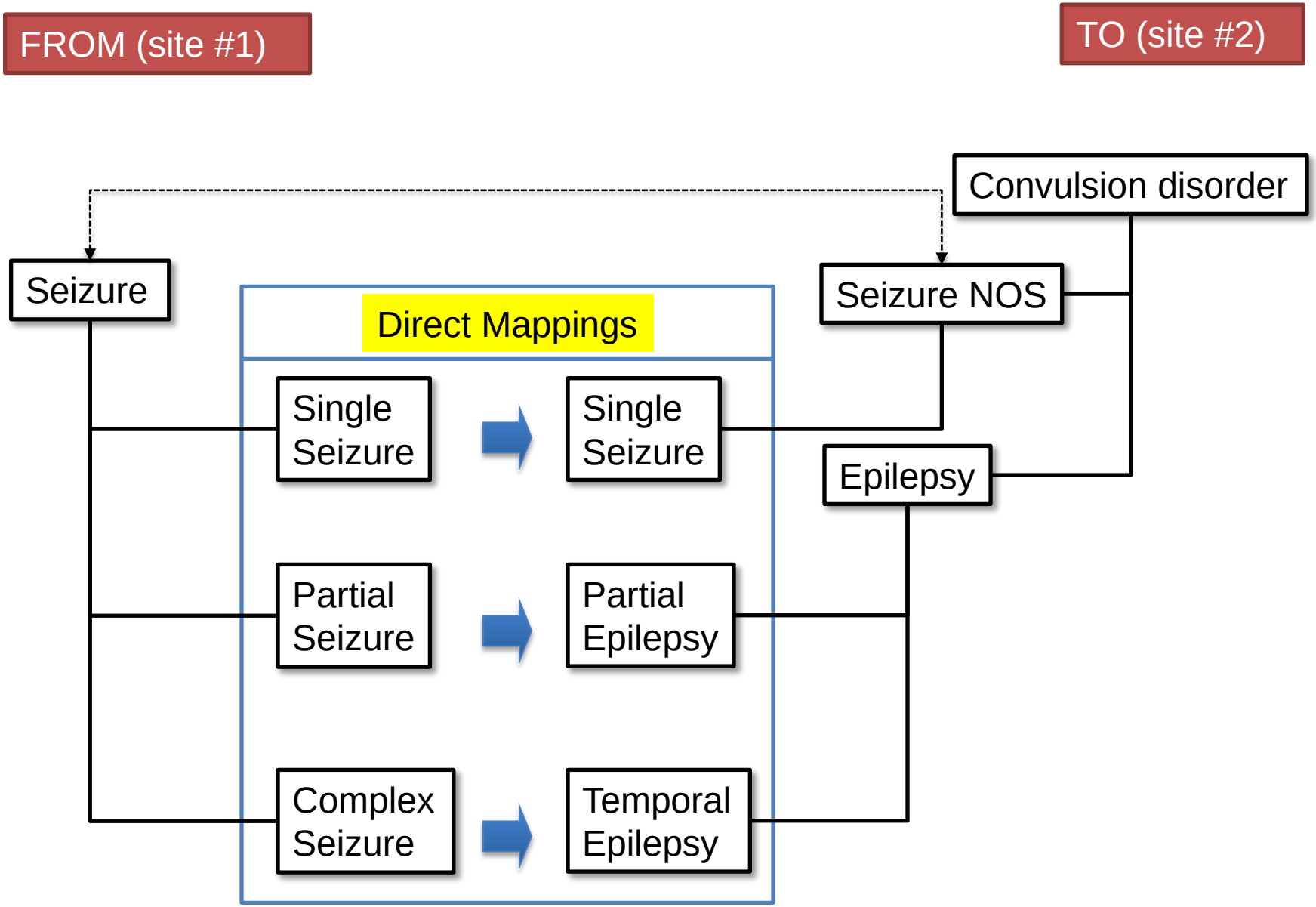


- Preliminary result:
  - Worked out the workflow for using mappings for query translation
  - Detailed discussions with the HOM and OpenMDR groups to define use-case and elicit requirements

- Future work:
  - Use BioPortal as the shared repository for inter terminology mappings
    - Tackle access, IP, performance, and institutional issues
  - Key features
    - ✓ Import outside mappings
    - ✓ Update mappings when versions change
    - ✓ Mechanism to curate mappings
    - Support proprietary curation and content

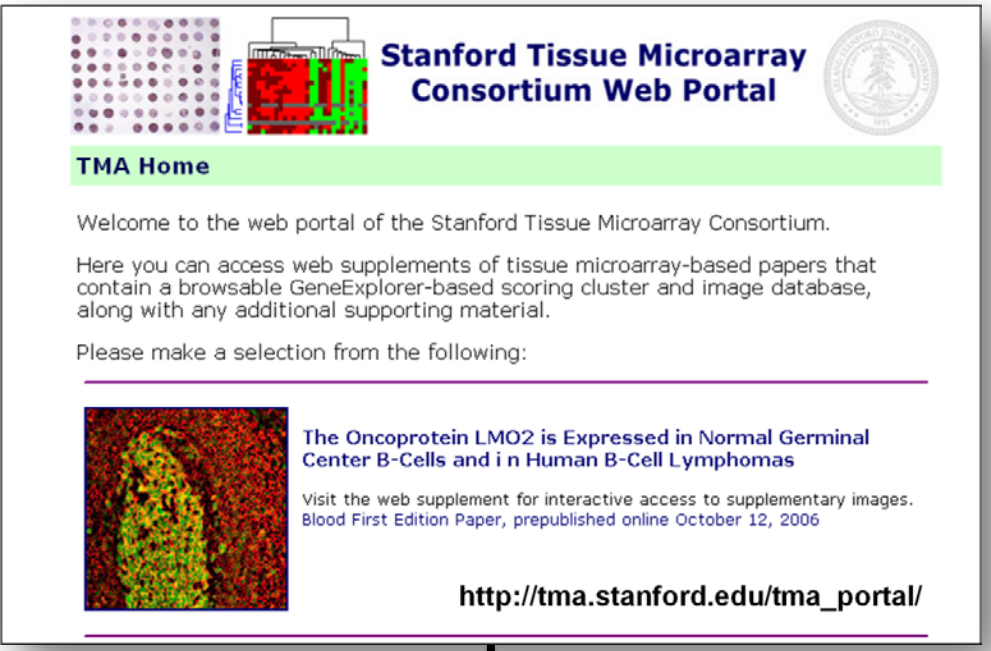
Using Mappings for query federation

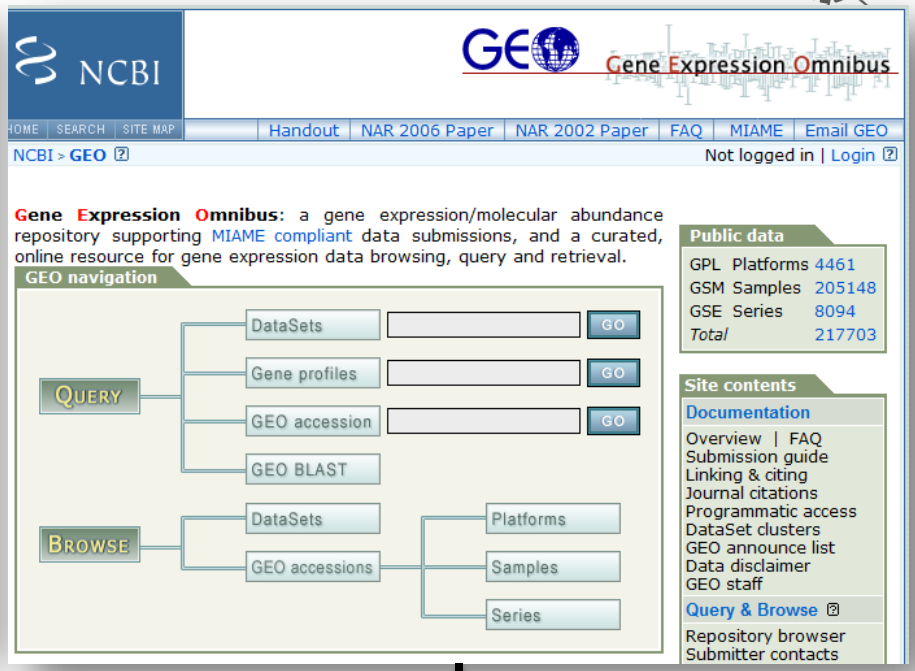




Use-cases based on automated annotation

Ontology based annotation

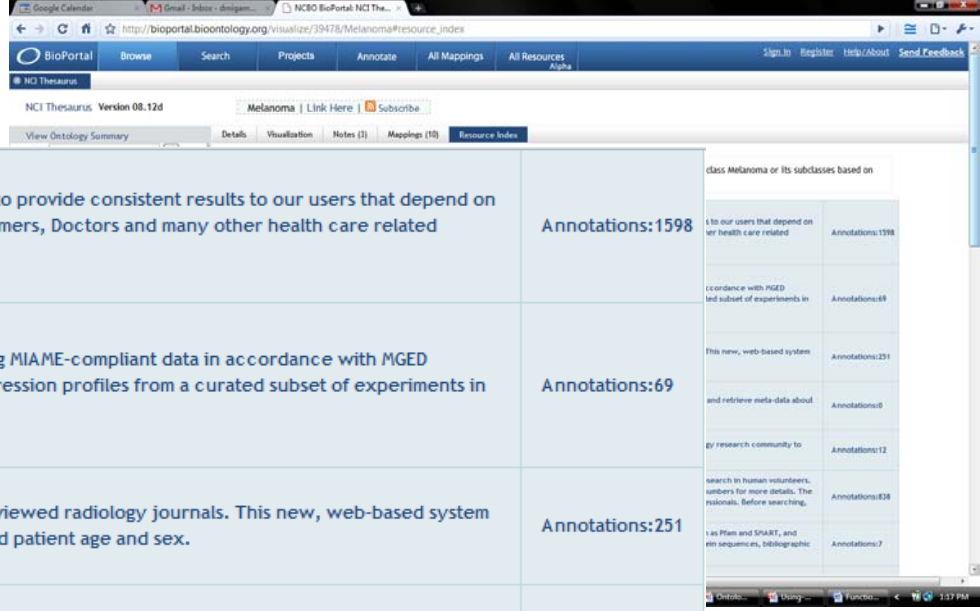




Disease	GEO samples	TMAD samples
Acute myeloid leukemia	366	3
Malignant melanoma	47	43
B-cell lymphoma	133	27
Prostate cancer	47	15
Renal carcinoma	34	185
Carcinoma squamous	105	175
Multiple myeloma	225	169
Clear cell carcinoma	34	63
Renal cell carcinoma	34	9
Breast carcinoma	3	1277
Hepatocellular carcinoma	80	163
Carcinoma lung	91	66
Cutaneous malignant melanoma	38	41
T-cell lymphoma	29	31
Lymphoblastic lymphoma	29	30
Uterine fibroid	10	19
Medulloblastoma	46	9
Clear cell sarcoma	35	8
Leiomyosarcoma	24	5
Mesothelioma	54	5
Kaposi's sarcoma	4	3
Cardiomyopathy	14	2
Dilated cardiomyopathy	14	2

20 diseases

Disease card



Adverse Event Reporting System Data	Adverse Event Reporting System (AERS) Databases are validated and normalized to provide consistent results to our users that depend on data such as this for research into Adverse Events reported to the FDA by Consumers, Doctors and many other health care related professionals.	Annotations:1598
ArrayExpress	ArrayExpress is a public repository for microarray data, which is aimed at storing MIAME-compliant data in accordance with MGED recommendations. The ArrayExpress Data Warehouse stores gene-indexed expression profiles from a curated subset of experiments in the repository.	Annotations:69
ARRS GoldMiner	ARRS GoldMiner provides instant access to images published in selected peer-reviewed radiology journals. This new, web-based system allows viewers to search for images by findings, anatomy, imaging technique, and patient age and sex.	Annotations:251
Biositemaps	Biositemaps represent a mechanism for computational biologists and bio-informaticians to openly broadcast and retrieve meta-data about biomedical data, tools and services (i.e., biomedical resources) over the Internet.	Annotations:0
caNanoLab	caNanoLab is a data sharing portal designed to facilitate information sharing in the biomedical nanotechnology research community to expedite and validate the use of nanotechnology in biomedicine.	Annotations:12
ClinicalTrials.gov	ClinicalTrials.gov provides regularly updated information about federally and privately supported clinical research in human volunteers. ClinicalTrials.gov gives you information about a trial's purpose, who may participate, locations, and phone numbers for more details. The information provided on ClinicalTrials.gov should be used in conjunction with advice from health care professionals. Before searching, you may want to learn more about clinical trials.	Annotations:838
Conserved Domain Database (CDD)	The Conserved Domain Database (CDD) contains protein domain models imported from outside sources, such as Pfam and SMART, and curated at NCBI. CDD contains over 12,000 such models and is linked to other NCBI databases, including protein sequences, bibliographic citations, and taxonomy.	Annotations:7
Database of Genotypes and Phenotypes	The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the results of studies that have investigated the interaction of genotype and phenotype. Such studies include genome-wide association studies, medical sequencing, molecular diagnostic assays, as well as association between genotype and non-clinical traits.	Annotations:0
DrugBank	DrugBank is offered to the public as a freely available resource. Use and re-distribution of the data, in whole or in part, for commercial purposes requires explicit permission of the authors and explicit acknowledgment of the source material (DrugBank) and the original publication.	Annotations:7
Gene Expression Omnibus DataSets	A gene expression/molecular abundance repository supporting MIAME compliant data submissions, and a curated, online resource for gene expression data browsing, query and retrieval.	Annotations:102

Linking annotations to data

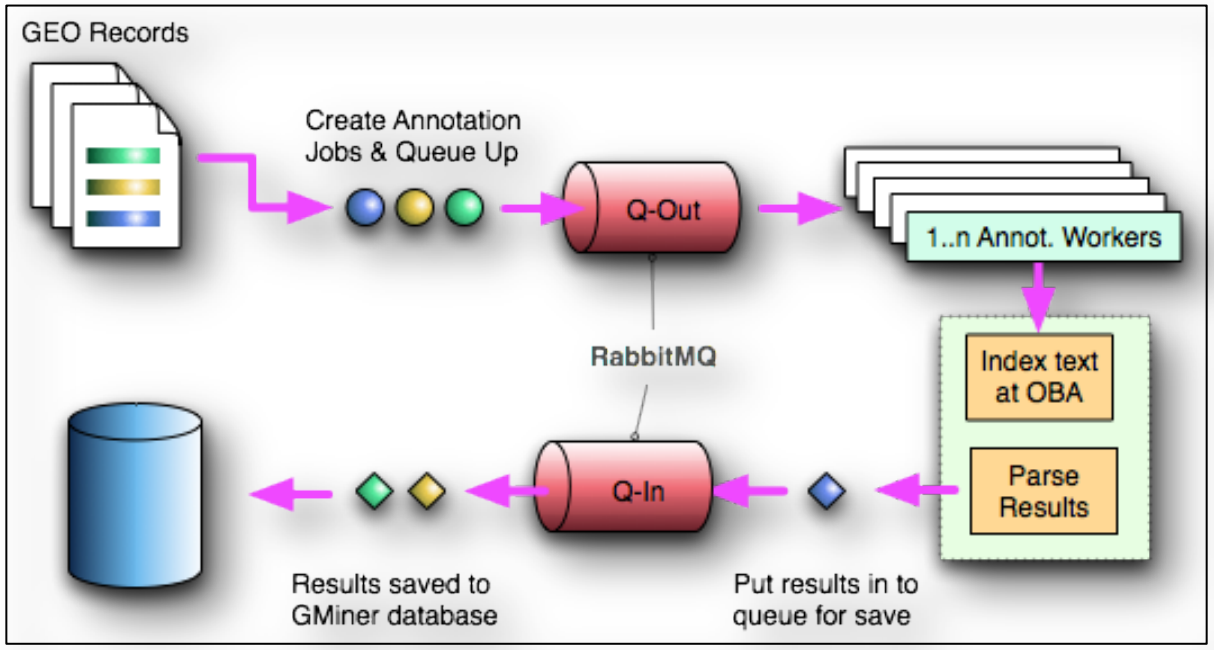
(by Simon Twigger)

Tm2d1
RGD1306410
Svs4
Hbb
Scgb2a1
Alb

1367550_a_at	417.4	P	0.000244
1367551_a_at	348.7	P	0.000244
1367552_at	9.6	A	0.366211
1367553_x_at	2293.3	P	0.000244
1367554_at	10	A	0.149658
1367555_at	14.4	A	0.111572

+

Sample GSM108325	
Status	Public on May 09, 2007
Title	Lewis replicate 1
Sample type	RNA
Source Name	Lewis kidney
Organism(s)	Rattus norvegicus
Characteristics	Lewis rat is wild type.
Extracted molecule	total RNA
Label	biotin



Annotations via Cloud

Anatomy terms

abdominal aorta accumbens nucleus adipose tissue adrenal gland adrenal gland medulla amygdala aorta artery articular cartilage blood blood vessel bone bone marrow bowel brain brainstem cardiac muscle tissue cardiovascular system cartilage caudate-putamen central nervous system cerebellum cerebral artery cerebral cortex cerebral hemisphere colon cranium cricoarytenoid dentate gyrus descending thoracic aorta diaphragm dorsal root ganglion duodenum endometrium epididymis extensor digitorum longus eye fat femur forebrain frontal cortex ganglion gastrocnemius gastrointestinal system gut heart heart endocardium heart left ventricle heart ventricle hindbrain hindlimb hindlimb muscle hip hippocampus hippocampus CA1 hypothalamus ileum interpeduncular nucleus intestine jejunum joint kidney knee lacrimal gland leg leg muscle ligament limb liver lung lymph mammary gland mesenteric artery midbrain muscle organ muscle tissue myocardium myometrium nerve nervous system ovary pancreas pancreatic islet pancreatic lymph node parietal cortex penis peripheral nervous system pineal gland pituitary gland putamen rectum rectus extraocular muscle retina saphenous artery seminiferous tubule epithelium serum set of skeletal muscles skeletal muscle tissue skeletal system skin small intestine smooth muscle tissue soleus spinal cord spinal nerve stomach stomach mucosa stomach wall subfornical organ substantia nigra substantia nigra pars compacta suprachiasmatic nucleus supraoptic nucleus telencephalon testis thoracic aorta tibialis cranialis uterus

Rat Strains

ACI ACI.COP-(D10Mgh8-D10Rat4)/ShuI ACI.COP-(D3Mgh16-D3Rat119)/ShuI ACI.COP-(D3Rat130-D3Rat114)/ShuI ACI.COP-(D6Rat80-D6Rat146)/ShuI BB BN congenic strain COP DA Eker F344 GH LE LE/Orl LEC LEW LOU NP NP/lusm P P/lusm PKD rat strain SBH SBN SD SHR SHRSP SR SR/JrHsd SS SS.SHR-(D2Rat61-D2Mco18)/Mco SS.SR-(D9Rat89-D9Mco27)/Mco SS/Jr WF WIST WKY Z

Annotations via Cloud

Anatomy terms

Rat Strains

SD SHR SS.SHR-(D2Rat61-D2Mco18)/Mco SS.SR-(D9Rat89-D9Mco27)/Mco SS/Jr

kidney AND SS

annotations reference 7 GEO records.

GEO accession	Description
GDS1545	Salt sensitive hypertension: kidney (RG-U34A)
GDS1546	Salt sensitive hypertension: kidney (RG-U34B)
GDS1547	Salt sensitive hypertension: kidney (RG-U34C)
GDS2710	Salt-sensitive hypertension model: kidney
GSE1775	Comparison of renal transcript levels in Dahl Salt-sensitive (S) rat with S.R congenic rat containing a BP QTL
GSE599	Dietary salt and renal function
GSE6208	Comparison of Renal Gene Expression from Dahl Salt-Sensitive (S) and the Renal Protective S.SHR(2) Congenic Strain

Tm2d1
RGD1306410
Svs4
Hbb
Scgb2a1
Alb

1367550_a_at	417.4	P	0.000244
1367551_a_at	348.7	P	0.000244
1367552_at	9.6	A	0.366211
1367553_x_at	2293.3	P	0.000244
1367554_at	10	A	0.149658
1367555_at	14.4	A	0.111572

+

Sample GSM108325

Status	Public on May 09, 2007
Title	Lewis replicate 1
Sample type	RNA
Source Name	Lewis kidney
Organism(s)	Rattus norvegicus
Characteristics	Lewis rat is wild type.
Extracted molecule	total RNA
Label	biotin

Hbb is_expressed_in rat kidney
Tm2d1 is_expressed_in rat kidney

Annotation Analytics

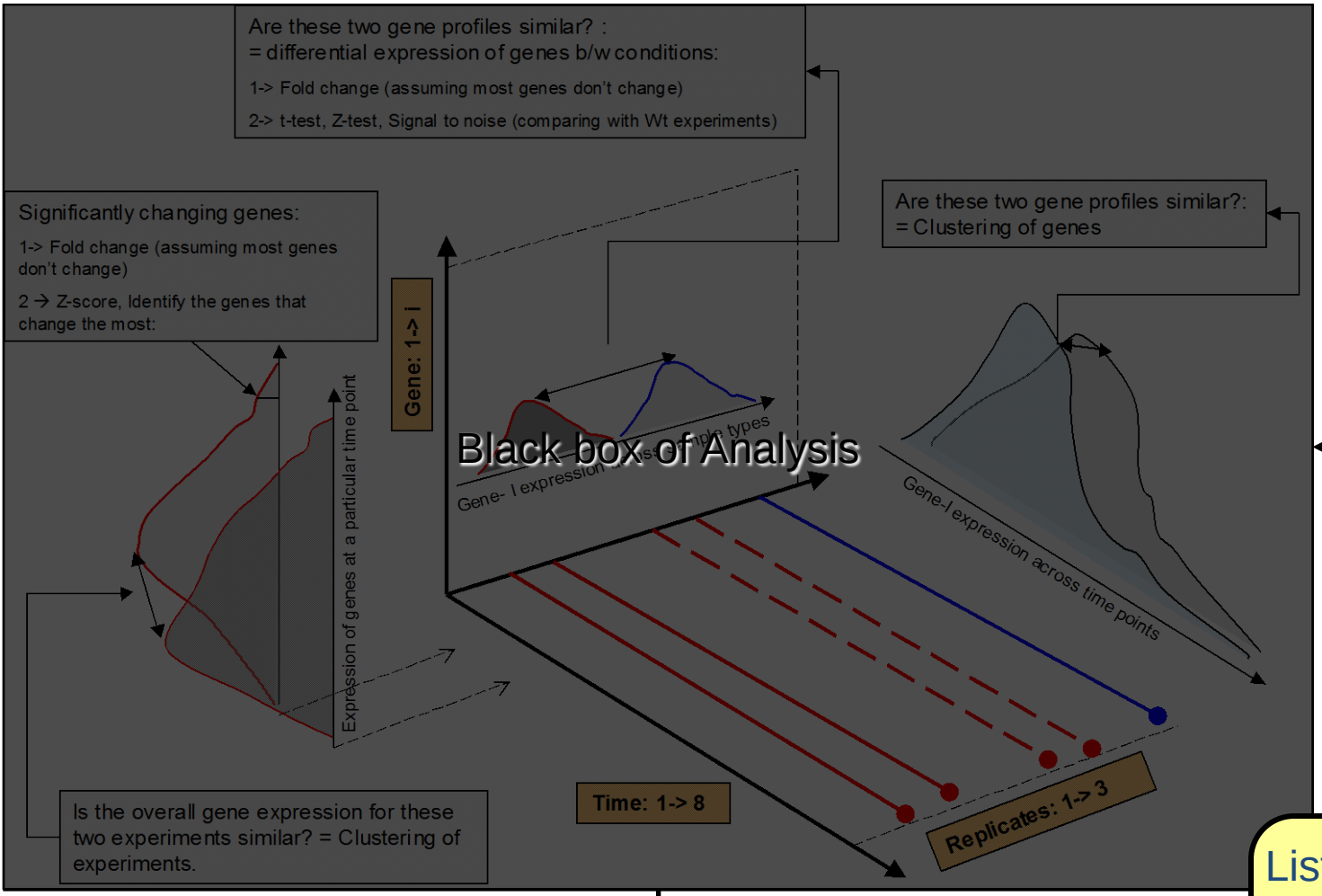
High throughput data

- “high throughput” is one of those fuzzy terms that is never really defined anywhere
- Genomics data is considered high throughput if:
  - You can not “look” at your data to interpret it
  - Generally speaking it means ~ 1000 or more genes and 20 or more samples.
- There are about 40 different high throughput genomics data generation technologies.
  - DNA, mRNA, proteins, metabolites … all can be measured

How do ontologies help?

- An ontology provides a *organizing framework* for creating “abstractions” of the high throughput data
- The simplest ontologies (terminologies, controlled vocabularies) provide the most bang-for-the-buck
  - Gene Ontology (GO) is the prime example
- More structured ontologies – such as those that represent pathways and more higher order biological concepts – still have to demonstrate real utility.

Analyzing Microarray data



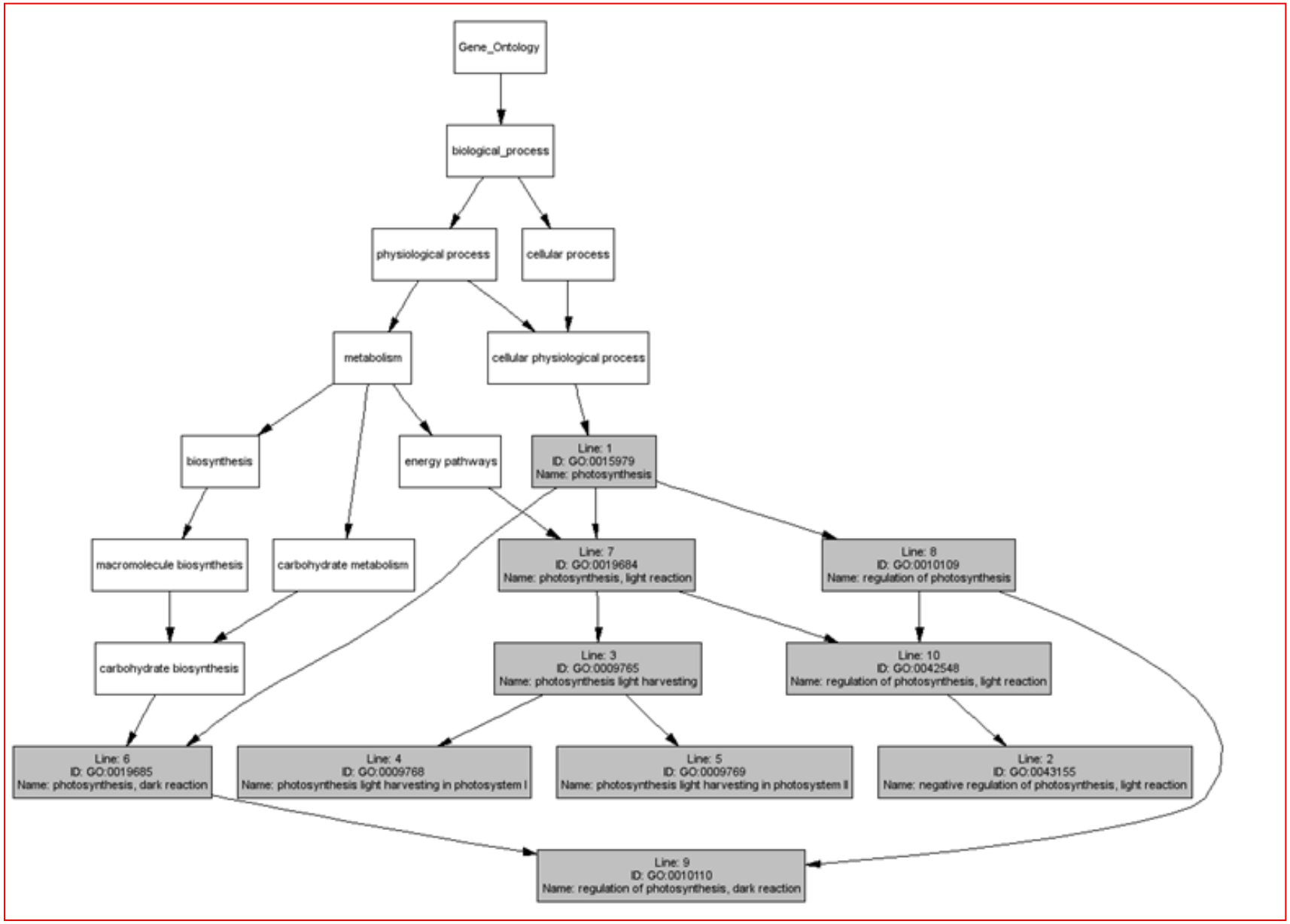
Raw Data:

Preprocessing:
- Normalization
- Flag 'bad' spots
- Handling duplicates
- Filtering
- Transformations

Lists of "Significantly changing" Genes.

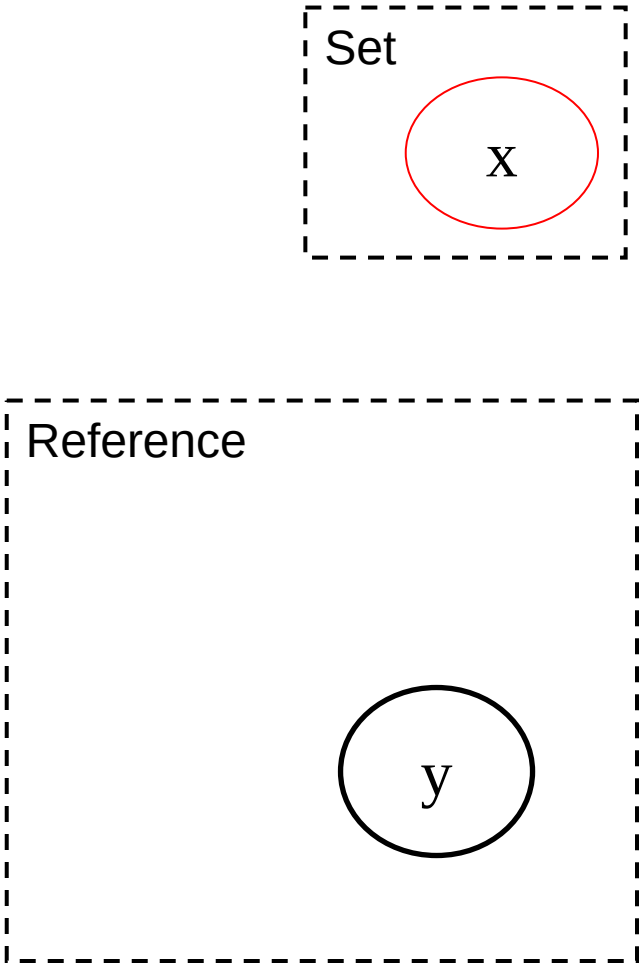
End up: **'Story telling'**

Gene Ontology (GO)



Generic GO based analysis routine

- Get annotations for each gene in list
- Count the occurrence (x) of each annotation term in gene list
- Count the occurrence (y) of that term in some reference set (whole genome?)
- P-value for how "surprising" is it to find x, given y.



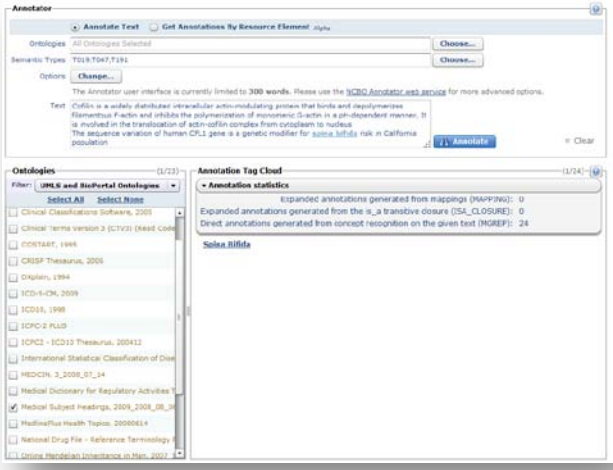
DIY Ontology Enrichment Analysis

Live Demo

Cfl1 ⟶ Cofilin is a widely distributed intracellular actin-modulating protein that binds and depolymerizes filamentous F-actin and inhibits the polymerization of monomeric G-actin in a pH-dependent manner. It is involved in the translocation of actin-cofilin complex from cytoplasm to nucleus. … The sequence variation of human CFL1 gene is a genetic modifier for spina bifida risk in California population

:

G-n ⟶ Some text …



http://rest.bioontology.org/obs/annotator

Cfl1 ⟶ spina bifida

:

G-n ⟶ Some disease condition

Cfl1 ⟶ spina bifida

:

G-n ⟶ Some disease condition

http://rest.bioontology.org/obs/rootpath/<ontologyid>/<conceptid>

Mutation enrichment

May have weak glycosidase activity towards glucuronylated steroids. However, it lacks essential active site Glu residues at positions 239 and 872, suggesting it may be inactive as a glycosidase in vivo. May be involved in the regulation of calcium and phosphorus homeostasis by inhibiting the synthesis of active vitamin D (By similarity). Essential factor for the specific interaction between FGF23 and FGFR1 (By similarity).

The Klotho peptide generated by cleavage of the membrane-bound isoform may be an anti-aging circulating hormone which would extend life span by inhibiting insulin/IGF1 signaling (By similarity).

	PHOS	METH	GLY	ACC	COIL	DIS	UBIQ	SHEET	CALM	CAT	HELIX	BUR	SIFT	STAB
Immune Function	-1.57	-1.31	-1.75	-2.44	-1.56	-3.00	-2.64	1.56	-2.19	-1.02	1.46	1.87	3.27	1.11
Neoplasm	1.15	-1.23	-2.22	-1.59	-1.21	-1.12	-1.77	-1.08	-1.62	1.09	1.50	1.54	2.69	1.12
Musculoskeletal	-3.60	-1.84	-8.02	-2.19	-1.15	1.53	-4.01	-1.11	1.40	1.21	1.41	1.80	3.35	1.11
Metabolic	-3.18	-3.44	-4.61	-2.55	-1.19	-1.31	-4.62	-1.13	-1.60	1.30	1.50	1.89	3.17	1.11
Dermatological	-1.56	-2.08	-3.66	-2.19	-1.19	1.15	-3.67	-1.25	-1.30	-1.05	1.56	1.80	3.18	1.05
Digestive System	-1.50	-1.40	-3.50	-1.76	-1.39	-1.87	-1.75	-1.22	1.35	1.24	1.86	1.63	2.80	1.13
Cardiovascular	-1.12	-1.49	-2.40	-1.52	-1.36	-1.03	-1.20	-1.25	2.58	1.07	1.83	1.50	2.91	1.13
Ophthalmological	-1.84	-3.27	-1.40	-2.04	-1.22	-1.85	-4.93	-1.15	1.51	1.05	1.57	1.75	2.96	1.10
Endocrine	-1.56	-2.08	-2.50	-2.00	-1.47	-1.91	-6.26	-1.45	1.44	1.30	2.06	1.73	3.16	1.13
Genitourinary	-2.59	-5.17	-3.46	-2.06	-1.32	-1.95	-2.60	1.08	2.20	1.17	1.56	1.75	2.95	1.16
Nervous System	-2.07	-1.02	-3.45	-1.95	-1.39	-2.18	-4.61	1.07	1.47	1.23	1.66	1.71	2.99	1.17
HGMD Disease Mutations	-2.12	-1.80	-3.58	-1.99	-1.28	-1.41	-2.29	1.05	1.40	1.16	1.51	1.73	3.09	1.11
Connective Tissue	-2.97	-2.97	0.00	-2.42	1.05	2.17	-8.95	-1.20	1.28	1.11	1.01	1.86	3.41	-1.00
Blood	-3.30	-1.46	-4.41	-2.17	-1.52	-2.11	0.00	1.02	-1.10	1.10	1.85	1.79	3.23	1.10
HGMD Functional Polymorphisms	-2.70	-1.35	-5.42	-1.42	-1.48	-2.24	2.58	-1.73	-1.13	-1.09	2.16	1.43	2.41	1.05
Blood Coagulation	-2.38	-2.38	-14.31	-1.92	-1.13	-3.94	-1.79	1.81	1.07	1.32	-1.55	1.70	3.14	1.17
Nutritional	-4.89	-3.26	-2.18	-2.66	-1.49	-7.11	0.00	-1.03	-1.09	1.33	1.85	1.91	3.06	1.17
Ear, Nose, Throat	-1.78	-1.19	0.00	-1.55	-1.35	-4.74	0.00	1.40	2.70	1.32	1.33	1.52	2.69	1.14
Lymphatic	0.00	0.00	0.00	-1.44	-1.21	-2.11	0.00	2.04	0.00	0.00	-1.67	1.45	3.20	1.00
Respiratory	-1.26	0.00	2.37	-1.49	-1.14	-1.01	2.36	1.07	3.33	-1.50	1.26	1.48	2.32	1.04
Psychiatric	1.22	2.56	1.10	-1.37	-1.19	1.30	0.00	-1.61	4.10	1.17	1.72	1.39	2.52	1.09
Developmental	1.43	2.14	-1.41	-1.38	1.01	1.34	-1.39	-1.20	4.00	-1.15	1.12	1.40	2.91	1.08
Cancer	2.21	2.26	-1.33	-1.05	-1.12	1.67	-1.11	-1.34	3.01	-1.36	1.47	1.06	1.95	1.00
Kinase	3.41	-2.10	-2.11	-1.05	1.04	1.19	-1.06	-1.64	1.71	-1.05	1.21	1.07	2.24	-1.07
HGMD Disease Polymorphisms	2.18	-1.53	1.52	-1.08	-1.06	1.32	0.00	-1.43	1.31	1.05	1.39	1.10	1.00	-1.03

Fold Depletion: >-5, -5, -3, -2, -1, 0, 1

Fold Enrichment: 2, 3, 4, 4<

Data mining: Drug, Disease, Gene relationships

Extraction of Conditional Probabilities of the Relationships Between Drugs, Diseases, and Genes from PubMed Guided by Relationships in PharmGKB

Martin Theobald, Ph.D[1], Nigam Shah, M.B.B.S., Ph.D[2], and Jeff Shrager, Ph.D[3]

Example:

p(salmeterol | Asthma, ADRB2) = 0.07
p(salbutamol | Asthma, ADRB2) = 0.16

At best these are pointers to hypotheses:

- Stronger biomarker?
- More reported side effects?
- Simple recency?
- *Many interpretations are possible!*

Disease Annotations – by Paea LePendu

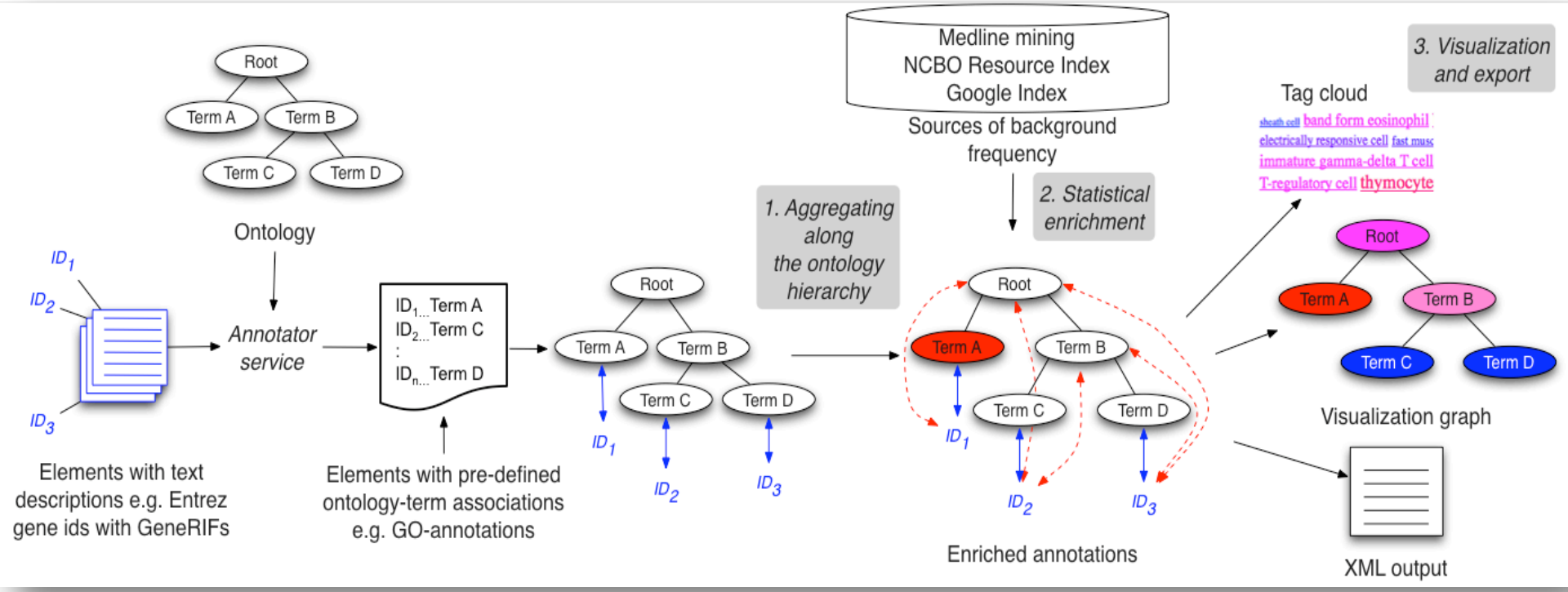
Cfl1 → Cofilin is a widely distributed intracellular actin-modulating protein that binds and depolymerizes filamentous F-actin and inhibits the polymerization of monomeric G-actin in a pH-dependent manner. It is involved in the translocation of actin-cofilin complex from cytoplasm to nucleus. … The sequence variation of human CFL1 gene is a genetic modifier for spina bifida risk in California population

:

:

G-n → Some text …

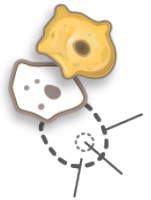
An Ontology Neutral analysis tool



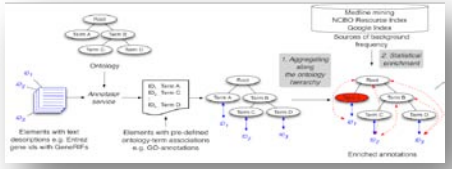
www.bioontology.org/wiki/index.php/Annotation_Summarizer

Sunday, 11/14, 3:30 – 5:00 PM, Session S07 in Lincoln West

Patient cohort analysis

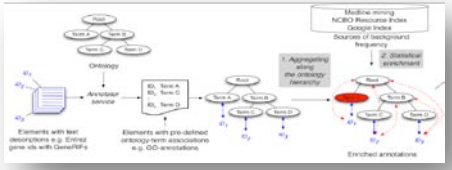


Extended criteria kidney transplant



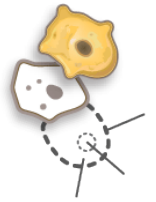
P (A | B, C …)

Standard criteria Kidney transplant

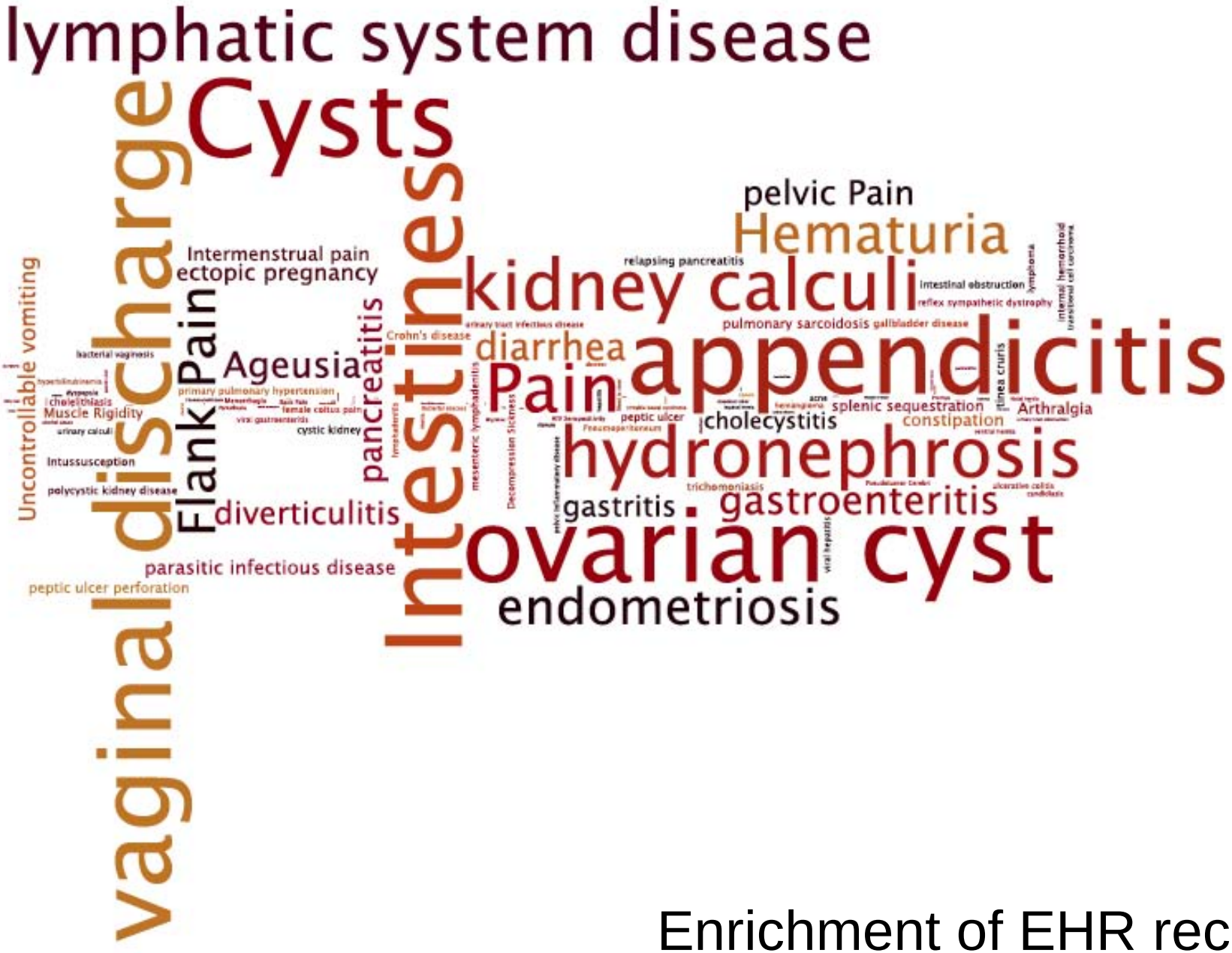


P (A | B, C …)

An “enrichment experiment” on EHRs



- Corpus of 90,000 EHR discharge summaries, progress reports, and other notes
- 233 Patients with discharge diagnosis, “Abdominal pain, not otherwise specified.”
- Identify other diagnoses (and their abstractions) that are “enriched” in this sample.



Enrichment of EHR records selected for “Abdominal Pain”

Data driven medicine

Data mining: Drug, Disease, Gene relationships

Extraction of Conditional Probabilities of the Relationships Between Drugs, Diseases, and Genes from PubMed Guided by Relationships in PharmGKB

Martin Theobald, Ph.D[1], Nigam Shah, M.B.B.S., Ph.D[2], and Jeff Shrager, Ph.D[3]

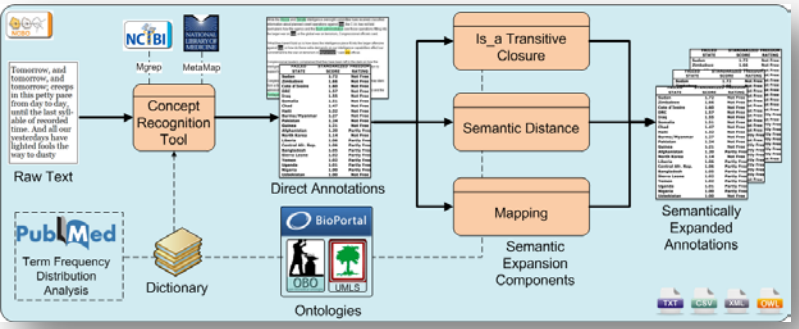
Example:

p(salmeterol | Asthma, ADRB2) = 0.07
p(salbutamol | Asthma, ADRB2) = 0.16

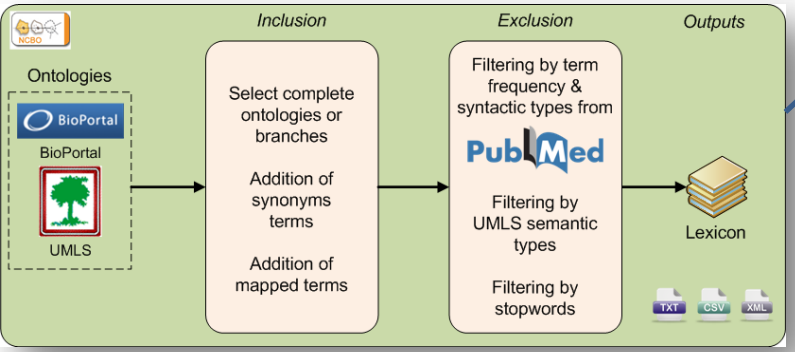
At best these are pointers to hypotheses:

- Stronger biomarker?
- More reported side effects?
- Simple recency?
- *Many interpretations are possible!*

Annotator



Lexicon Builder



Tuesday, Nov 16th, 3:30 to 5:00 PM
S76: Use of Knowledge Sources
Biomedicine and Discovery

Outcome of interest

ID-1	fever	pain	catheter	prostate	mass	diabetes	bleeding
ID-2	no pain	swelling	abdomen	burning	kidney	stone	724.0
:							
:							
:							
:							
:							
:							
:							
:							
ID-n	Code-1	Code-2	Code-8	Code-6	Code-3	Code-5	Code-2

ID-1	fever	pain	catheter	prostate	mass	diabetes	bleeding
ID-2	no pain	swelling	abdomen	burning	kidney	stone	724.0
:							
:							
:							
:							
:							
:							
:							
:							
ID-n	Code-1	Code-2	Code-8	Code-6	Code-3	Code-5	Code-2

Outcome of interest

Example:

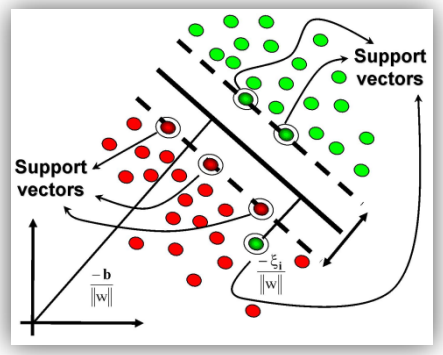
p(salmeterol | Asthma, ADRB2) = 0.07

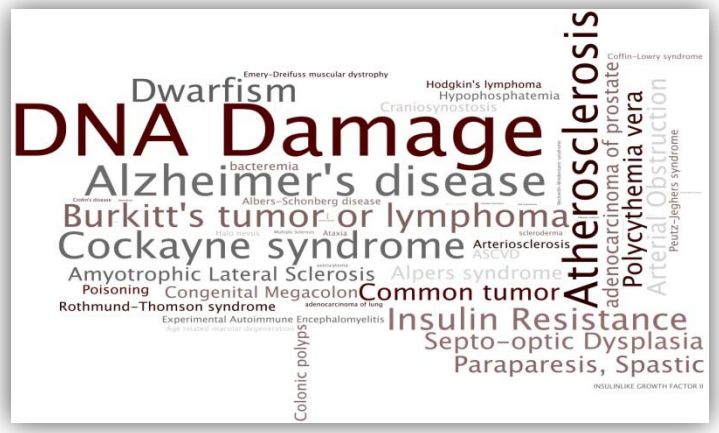
p(salbutamol | Asthma, ADRB2) = 0.16

p(readmission|Code-1, Code-8) = 0.70

These are pointers to hypotheses:

- Stronger biomarker?
- More reported side effects?
- High risk of readmission?



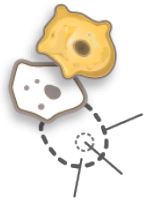


Learning and Prediction

Classification

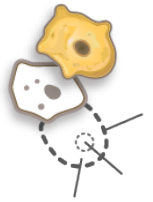
Summarization

Research questions

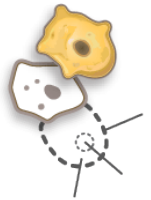


- Is it possible to perform “big-data” mining on medical records?
- Disease management
- Data-driven guideline design
- Risk and outcome prediction
  - E.g. 30 day readmission for patients covered by Medicare cost $17.4 billion (2004 data in Jencks et al NEJM, April 2, 2009)

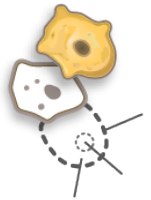
NCBO related talks



- Sunday, Nov 14th, 3:30 PM to 5:00 PM
  - S07: Making the Most of Ontologies
- Monday, Nov 15th, 3:30 PM to 5:00 PM
  - S39: Informatics Tools to Support Ontology Development
- Tuesday, Nov 16th, 1:45 PM to 3:15 PM
  - S62: Managing Terminologies to Realize Clinical Benefits
- Tuesday, Nov 16th, 3:30 PM to 5:00 PM
  - S76: Use of Knowledge Sources Biomedicine and Discovery



THE END

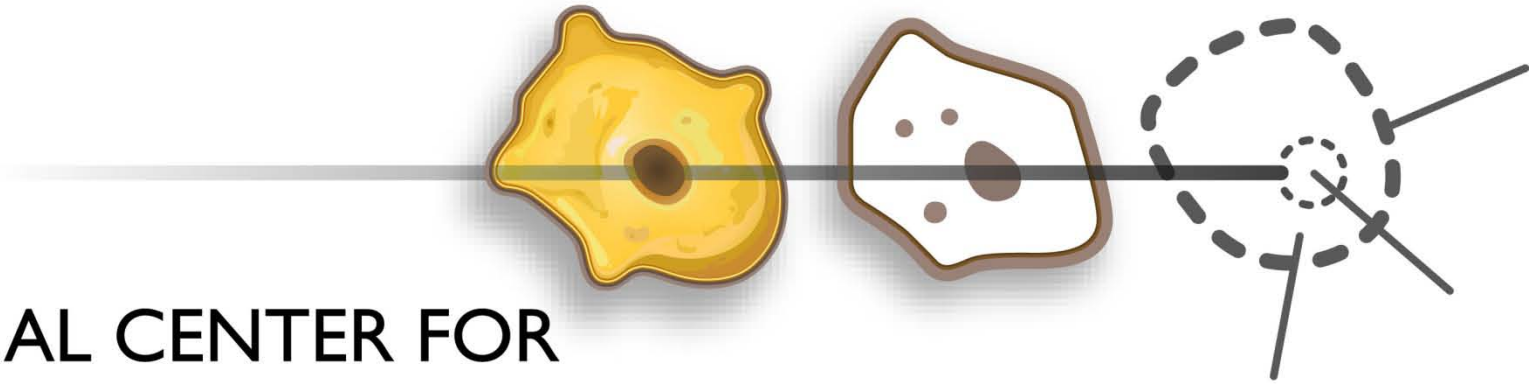


Credits

Mark Musen, PI

The team @
www.bioontology.org/project-team

NIH Roadmap grant U54 HG004028



NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY