

Annotation Analysis for Testing Drug Safety Signals

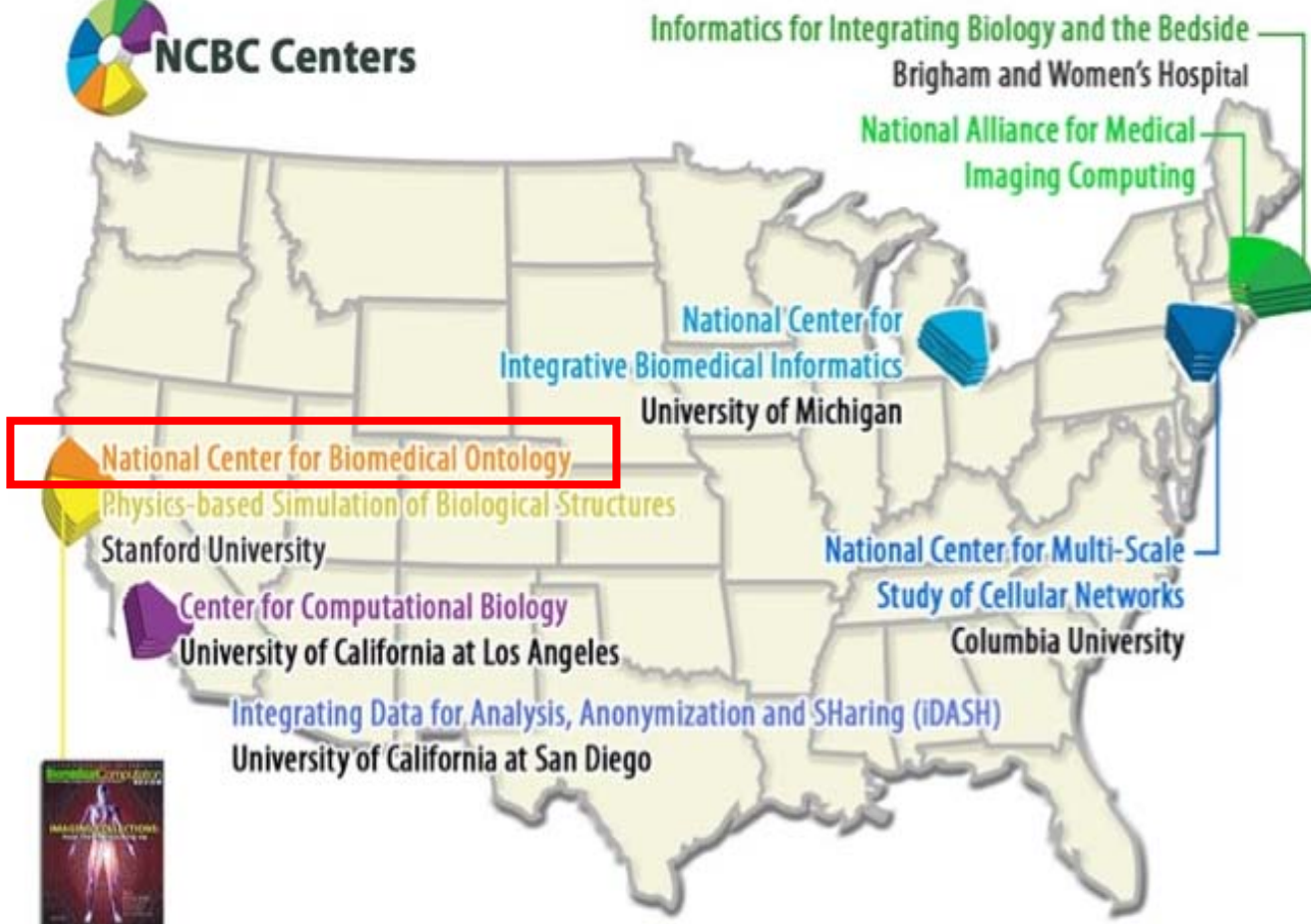
Trish Whetzel

whetzel@stanford.edu

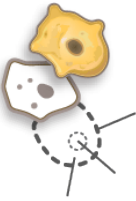




NCBC Centers



National Centers for Biomedical Computing
(<http://www.ncbcs.org>)



National Center for Biomedical Ontology

- Mission
 - To create software for the application of ontologies in biomedical science and clinical care
- NCBO Partners
 - Stanford University - Dr. Mark A. Musen
 - Mayo Clinic - Dr. Christopher G. Chute
 - University at Buffalo - Dr. Barry Smith
 - University of Victoria - Dr. Margaret-Anne Storey



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**

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.



www.bioontology.org

National Center for Biomedical Ontology

Community ▶

[Learning About Ontologies ▶](#)

[Dissemination & Training ▶](#)

[NCBO Collaborations ▶](#)

[Forum, Blog Publications ▶](#)

Technology ▶

[Ontology Library ▶](#)

[Go to BioPortal ▶](#)

[Data Annotation ▶](#)

[Go to Annotator ▶](#)

[Ontology Development ▶](#)

[Data Access Using Ontologies ▶](#)

[Go to Resource Index ▶](#)



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.

FILTER BY CATEGORY

All Categories ▾

FILTER BY GROUP

All Groups ▾

FILTER BY TEXT

[Submit New Ontology](#) [Subscribe to all updates](#)

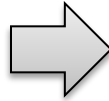
ONTOLOGY NAME ▲	VISIBILITY	TERMS	NOTES	REVIEWS	PROJECTS	UPLOADED	AUTHOR
ABA Adult Mouse Brain (ABA)	Public	913	0	0	2	08/08/2009	Allen Institute for Brain Science
Adverse Event Ontology (AEO)	Public	431	0	0	0	04/13/2011	Yongqun "Oliver" He
Adverse Event Reporting ontology (AERO)	Public	240	0	0	1	04/09/2011	Melanie Courtot
African Traditional Medicine (ATMO)	Public	223	2	2	2	06/28/2009	Ghislain Atemezang
AI/RHEUM (AIR)	Public	681	0	0	0	02/05/2010	May Cheh
Amino Acid (amino-acid)	Public	46	5	0	3	07/02/2010	Nick Drummond, Georgina Moulton, Robert Stevens, Phil Lord
Amphibian gross anatomy (AAO)	Public	700	0	0	3	02/15/2011	David Blackburn
Amphibian taxonomy (ATO)	Public	6,135	0	0	1	11/02/2009	AmphiAnat list
Anatomical Entity Ontology (AEO)	Public	137	0	0	0	02/15/2011	EMAP Administrators

<http://bioportal.bioontology.org>

<http://rest.bioontology.org>

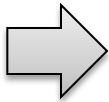
Ontology
Services

Views



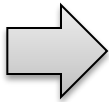
- Search
- Traverse
- Comment
- Download

Mapping
Services



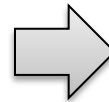
- Create
- Download
- Upload

Widgets



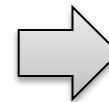
- Tree-view
- Auto-complete
- Graph-view

Annotation



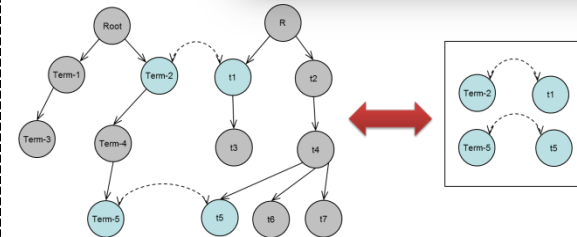
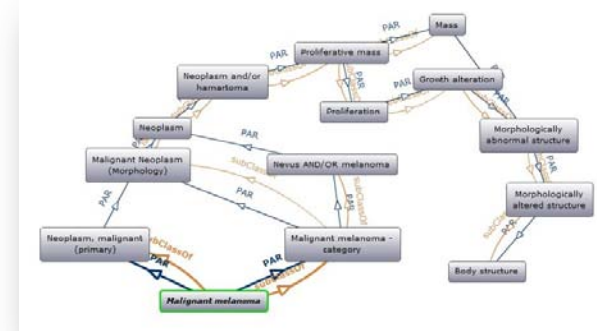
Term recognition

Data Access



Fetch "data"
annotated with a
given term

<http://bioportal.bioontology.org>

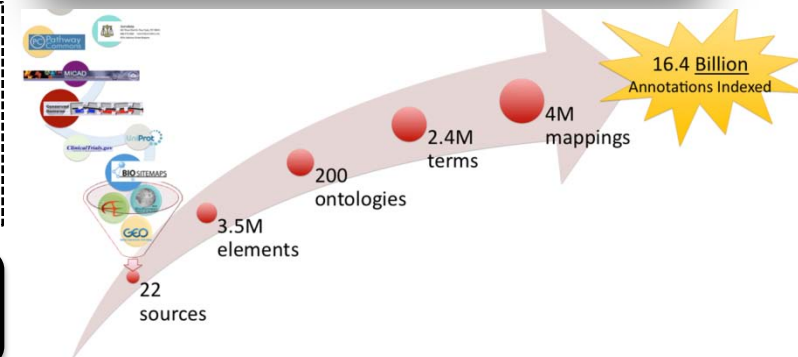


Jump To:

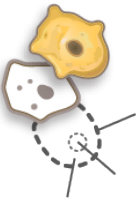
Legend

- Malignant **melanoma** (synonym)
- Amelanotic **melanoma** (preferred name)
- Excision of **melanoma** (preferred name)
- Melanoma** in situ (preferred name)
- Melanoma** vaccine (preferred name)

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 Web service



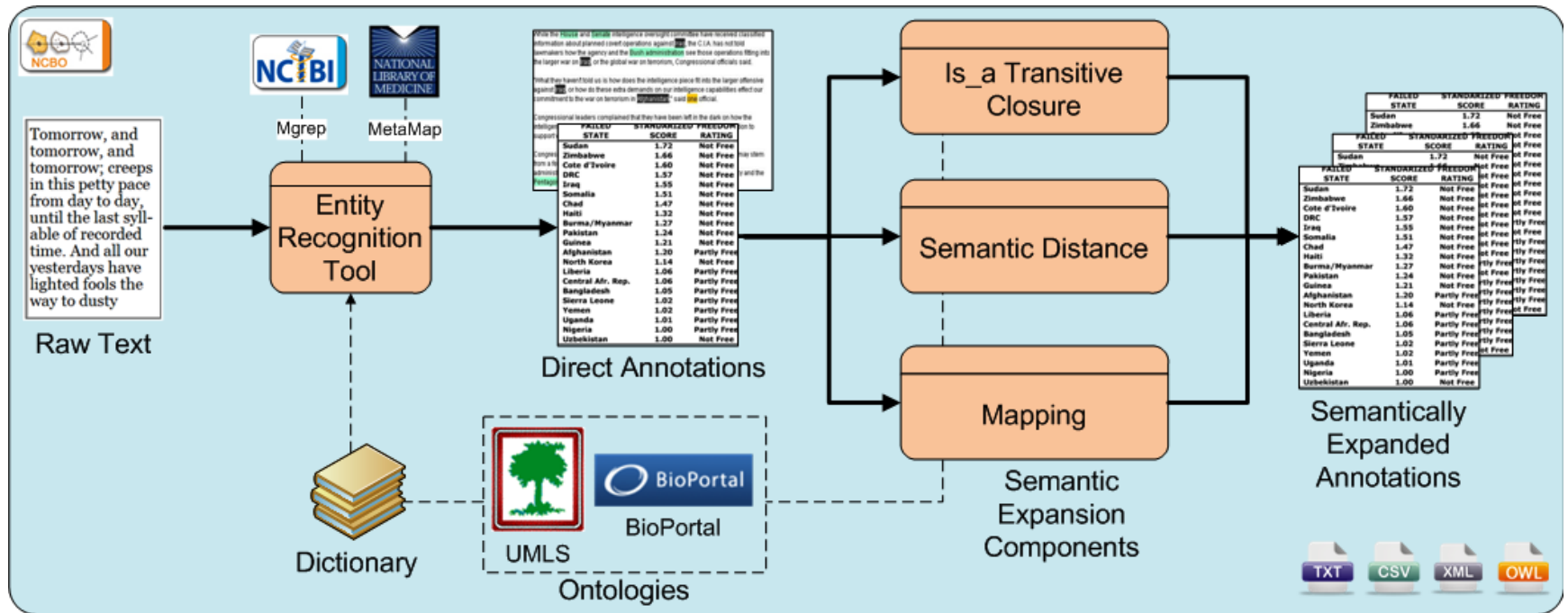
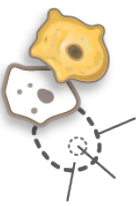
Process textual metadata to automatically tag text with as many ontology terms as possible.

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. <i>Physiol Genomics</i> 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



Resource Index

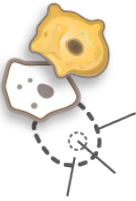


NCBO is building a system for automated ontology-based annotation and indexing of biomedical data. We process the textual metadata of diverse elements of biomedical resources such as gene expression data sets, descriptions of radiology images, clinical-trial reports, and PubMed abstracts to annotate and index them with terms from appropriate ontologies. Use the interface below to search the resulting index of annotations and to identify biomedical data resources annotated with particular ontology terms. Try searching for "Melanoma" and use NCI Thesaurus as your ontology.

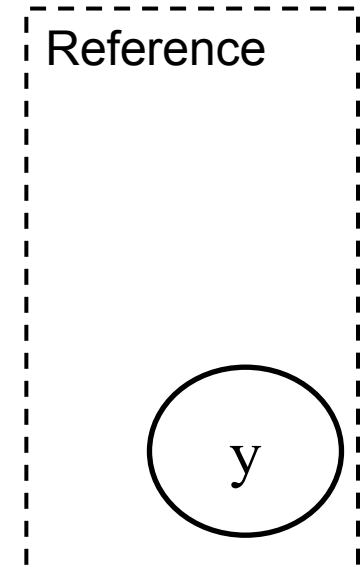
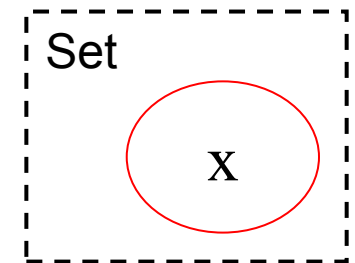
BioPortal Resource Search

[View Demo!](#)
Type in a search term[Search](#)[Clear](#)[1155](#)  **ARRS GoldMiner**[15190](#)  **ArrayExpress**[98304](#)  **ClinicalTrials.gov**[251](#)  **Database of Genotypes and Phenotypes**[21969](#)  **Gene Expression Omnibus DataSets**[21242](#)  **Online Mendelian Inheritance in Man**[832](#)  **PharmGKB [Disease]**[992](#)  **PharmGKB [Gene]**[87858](#)  **Reactome**[18710](#)  **Stanford Microarray Database**[1494](#)  **WikiPathways**[1172881](#) **Adverse Event Reporting System Data**[1653](#)  **Biositemaps**[40809](#)  **Conserved Domain Database (CDD)**[4774](#)  **DrugBank**[823](#)  **MICAD**[923](#)  **Pathway Commons**[1634](#)  **PharmGKB [Drug]**[110481](#)  **PubChem**[1033651](#)  **ResearchCrossroads**[18344](#)  **UniProt KB**[845](#)  **caNanoLab**

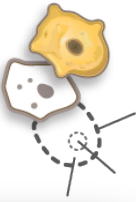
Generic GO based analysis routine



- 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.



Enrichment Analysis with the DO



<http://www.geneontology.org/GO.down>

Annotation Details and Downloads
Filtered files

{ERCC6, PARP1} ⇔ PMID:16107709

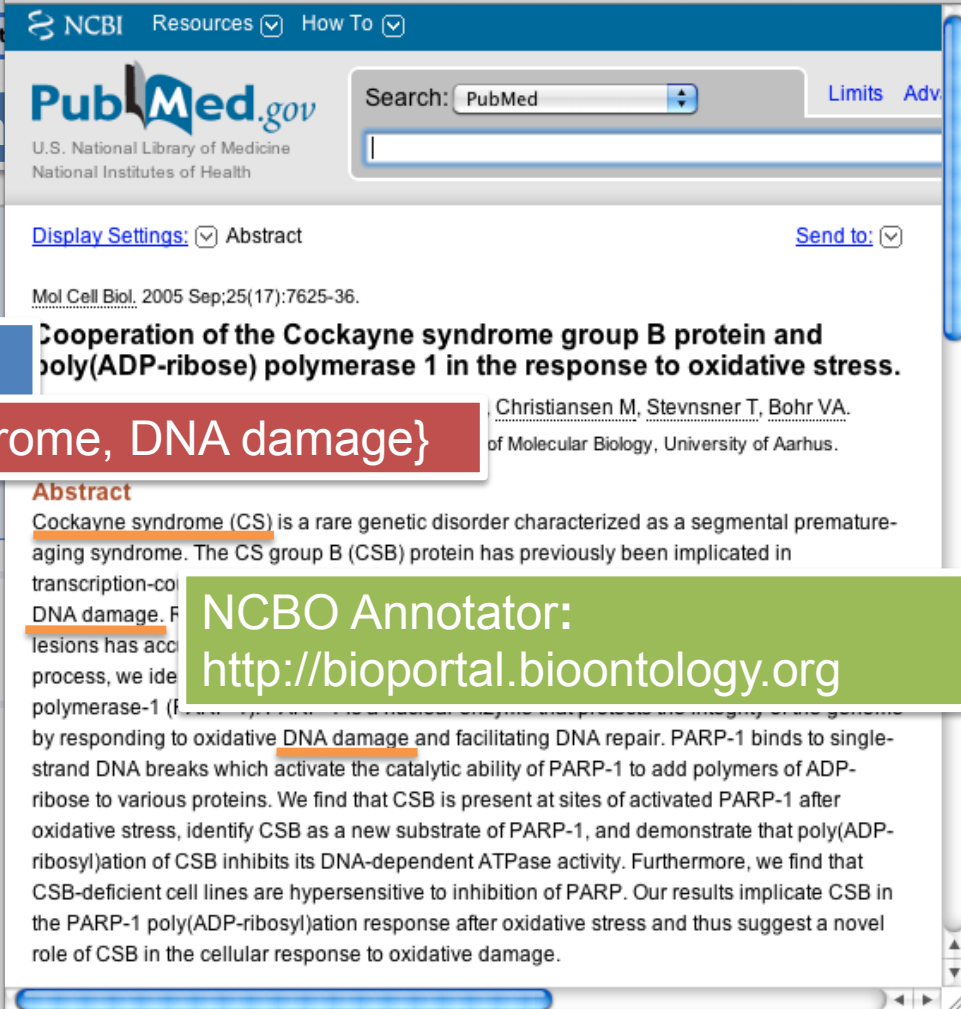
{ERCC6, PARP1} ⇔ {Cockayne syndrome, DNA damage}

Annotation Details and Downloads

Homo sapiens 18216
GO Annotations @ EBI

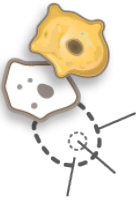
ERCC6
ERCC6
PARP1
ERCC6
PARP1

www.ncbi.nlm.nih.gov/pubmed/16107709

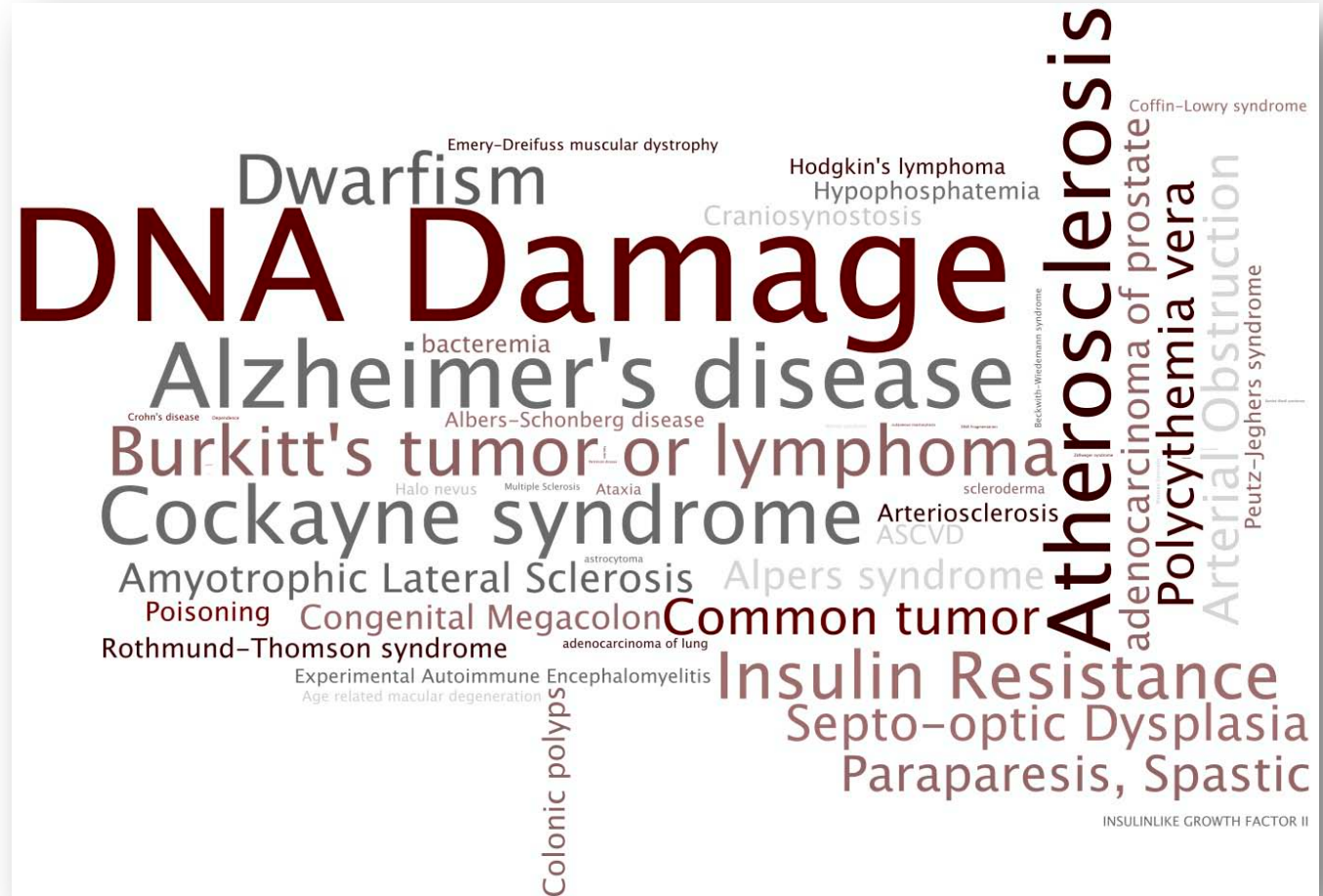


NCBO Annotator:
<http://bioportal.bioontology.org>

Profiling a set of Aging genes



Aging-related genes (261) – <http://genomics.senescence.info/genes/>





This text has been removed from the original slide show presentation as per our data usage agreements with the providers of the patient data.

This text originally contained a sample of a de-identified patient note.



Ontologies

All Ontologies Selected

Choose...

Semantic Types

All Semantic Types Selected

Choose...

Options

Change...

The Annotator user interface is currently limited to 300 words. Please use the [NCBO Annotator web service](#) for more advanced features.

Text

REASON FOR HOSPITALIZATION: X is a 48-year-old man with left-sided neck pain and radiation down to the left side of his shoulder and upper arm with numbness of the thumb, index and middle fingers at the tips, more on the left, and ongoing for the past three months. The patient had an MRI scan of the cervical spine done in early MONTH, which revealed cervical

Annotate

Annotation Term List

Annotation Tag Cloud

Terms

Filter: type filter text

Select All Select None

☒ (Hypertensive disease) or (hypertension) (2)

☒ (Pain) or (C/O: [ache] or [pain]) (1)

☒ (Spondyloses: [cervical] or [lumbar] or [sac

☒ Aciphex (1)

☒ Admission (4)

☒ Admission diagnosis (2)

☒ Advair (1)

☒ Advair Diskus (1)

☒ Adverse Event Associated with Pain (1)

Annotations

Filter: type filter text (filters by Term or Ontology)

Term: (Hypertensive disease) or (hypertension)

Ontology: SNOMED Clinical Terms

Context: PAST MEDICAL HISTORY: Hypertension.

Term: (Hypertensive disease) or (hypertension)

Ontology: SNOMED Clinical Terms

Context: PAST MEDICAL HISTORY: Hypertension.

Term: (Pain) or (C/O: [ache] or [pain])

Ontology: SNOMED Clinical Terms

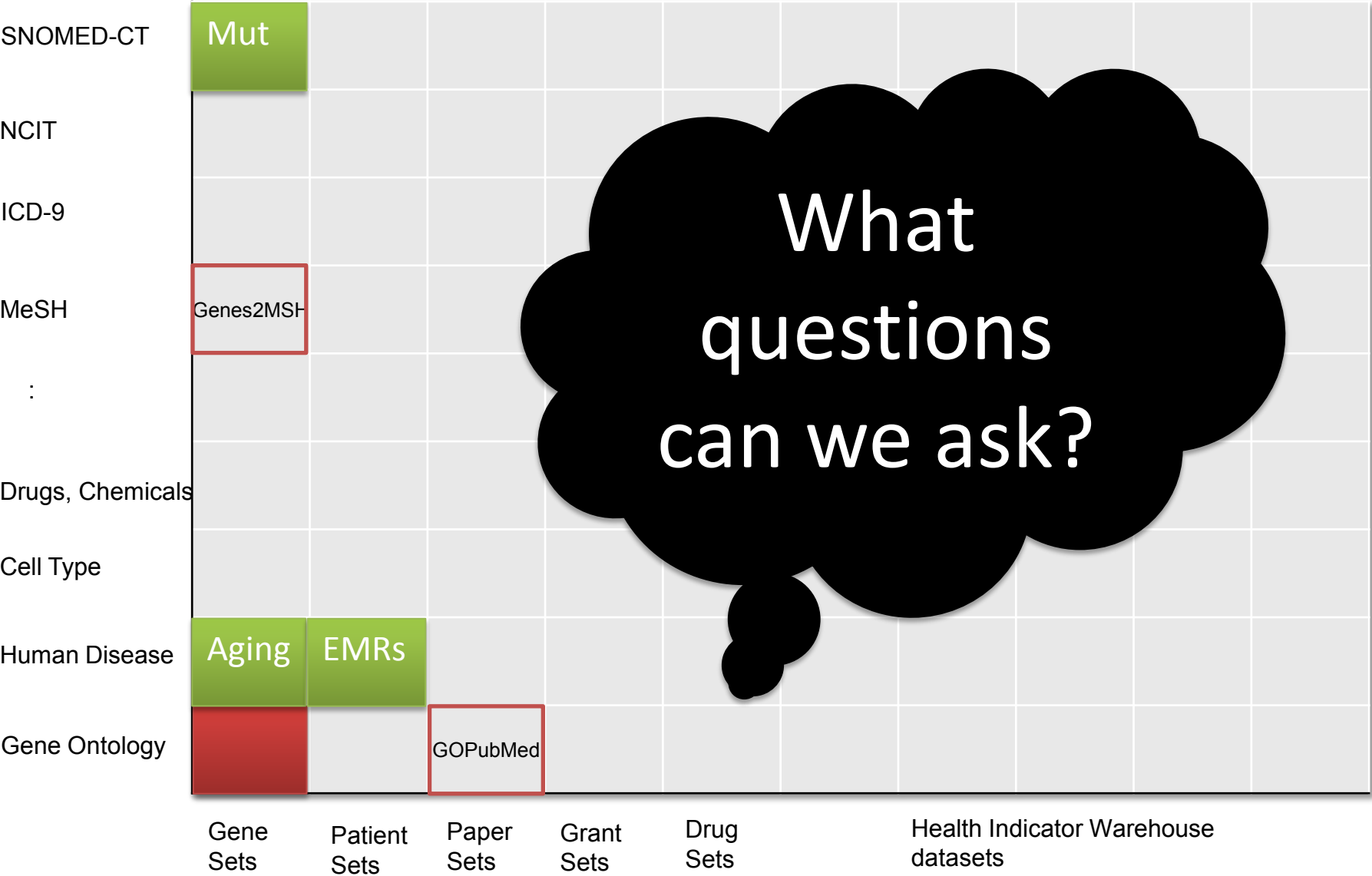
Context: PROCEDURES PERFORMED: C5-C6 anterior cervical diskectomy and fusion v instrumentation, performed by Y

(Abdominal pain, unspecified)



Patient records processed from U. Pittsburgh NLP Repository with IRB approval.

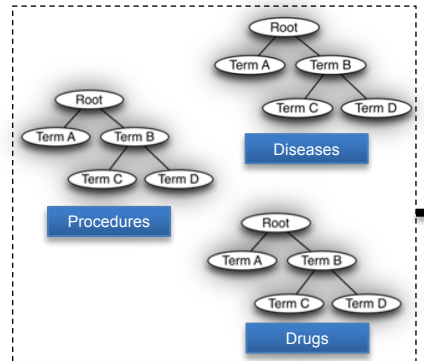
Annotation Analytics Landscape



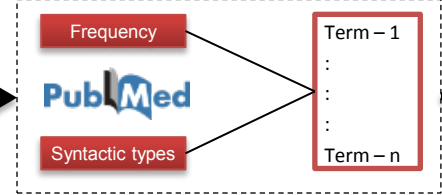
Generation of tagged data



BioPortal – knowledge graph



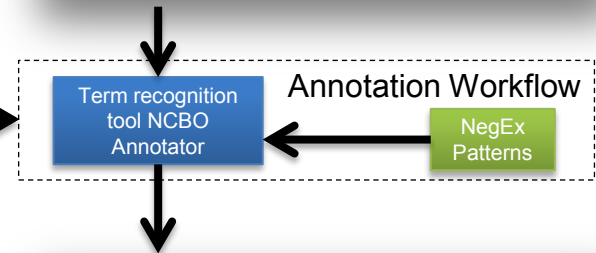
Creating clean lexicons



T/SICU Nursing Admission Note:

This is a 31 year old male s/p seizure on ladder with resulting fall 15-20 feet on [**09-17**] now presenting to the T/SICU post surgical repair of multiple facial fractures, right mandibular fracture, and left distal radius fracture. He needs to remain intubated for 48 hours post-op. His past medical history is significant only for seizure disorder, and his only medication is depakote. He has no known allergies.

Text clinical note



Nursing Admission Note:

is a year old male seizure 20 feet presenting post surgical repair multiple facial fractures, right mandibular fracture, left distal radius fracture. needs 48 hours post-op. past medical history significant seizure disorder, medication depakote. no known allergies.

Terms Recognized

NegEx Rules – Negation detection

Nursing Admission Note:

is a year old male seizure 20 feet presenting post surgical repair multiple facial fractures, right mandibular fracture, left distal radius fracture. needs 48 hours post-op. past medical history significant seizure disorder, medication depakote. no known allergies.

Negation detection

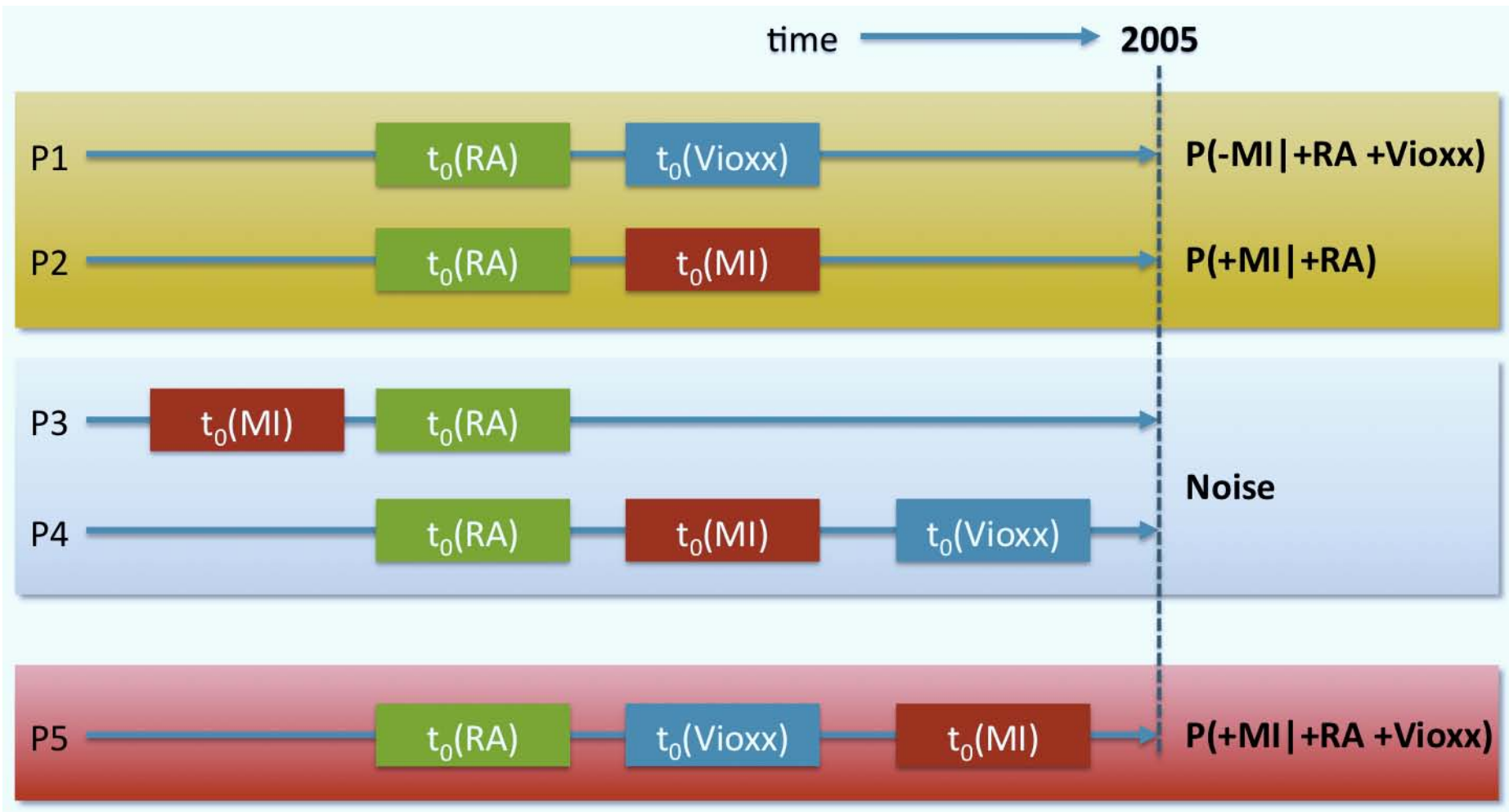
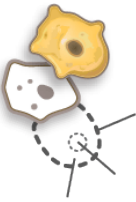
Further Analysis

Cohort of Interest

P1	ICD9		ICD9		ICD9	ICD9		ICD9	ICD9
P1	T1, T2, no T4		T5, T4, T3		T4, T3, T1	T8, T9, T4		T6, T8, T10	T1, T2, no T4
P2									
P2									
P3									
P3									
.									
.									

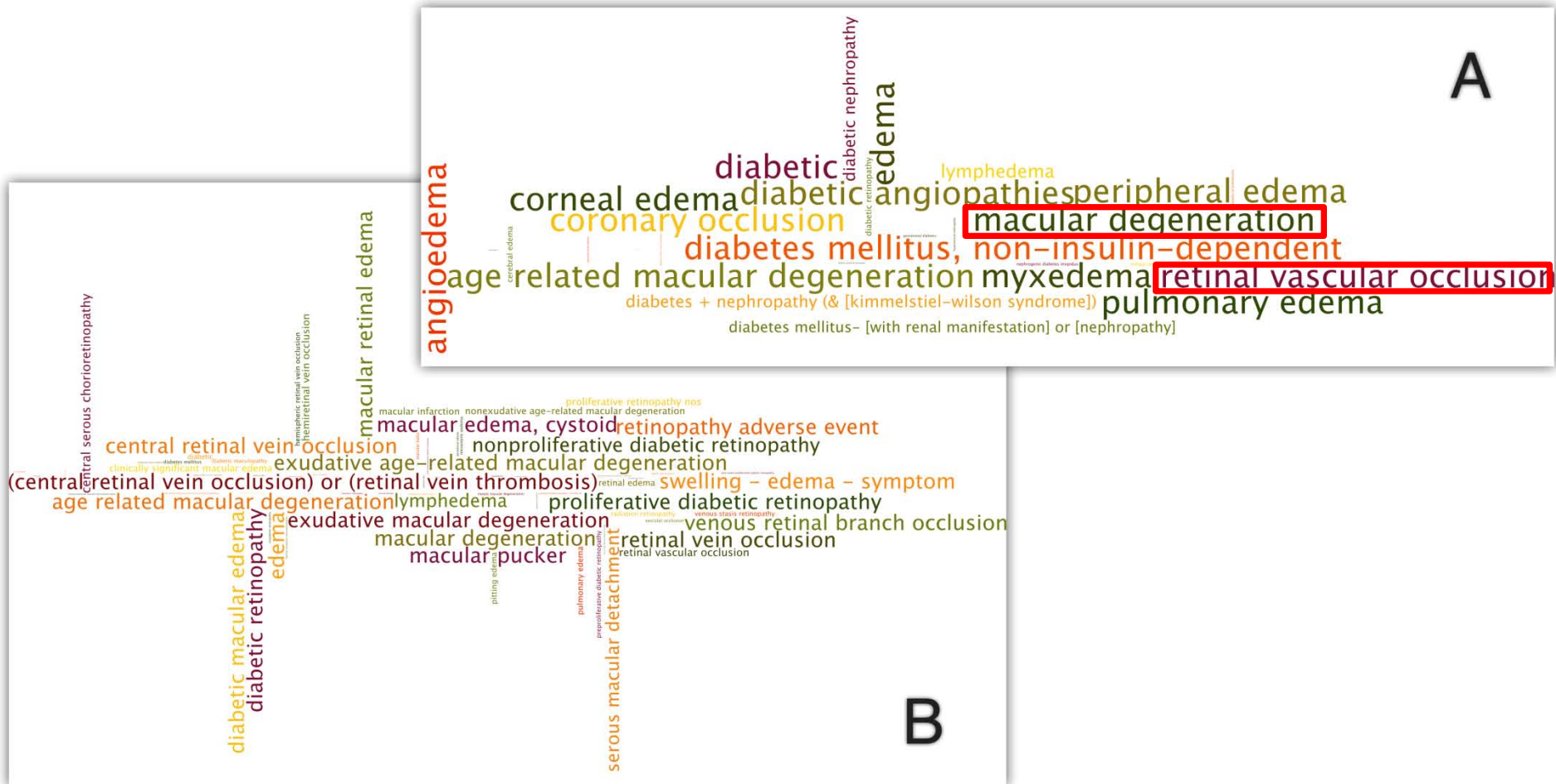
Terms form a temporal series of tags →

Adverse drug events

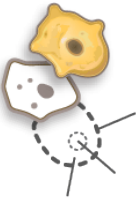


ROR of 2.058, CI of [1.804, 2.349]
 PRR of 1.828, CI of [1.645, 2.032]
 The uncorrected χ^2 statistic has p-value $< 10^{-7}$.

ROR=1.524, CI=[0.872, 2.666]
 PRR=1.508, CI=[0.8768, 2.594]
 χ^2 p-value=0.06816.



Recent Developments



From Lexicons to Pattern Recognition:

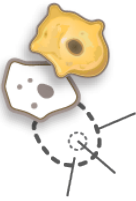
- Smoking status patterns
- Medication dosage patterns
- Other expert-defined patterns (GUI)

Validation:

- 93% recall, 53% precision for drug mentions (RxNORM)

Advanced Analysis Methods:

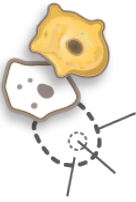
- Propensity score adjustment methods
- Mapping and normalization techniques



END

Annotator enables mining clinical text.

Acknowledgements



- Paea LePendur
- Stephen Racunas
- Srinivasan Iyer
- Yi Liu
- Cedrick Fairon
- Nigam Shah – nigam@stanford.edu