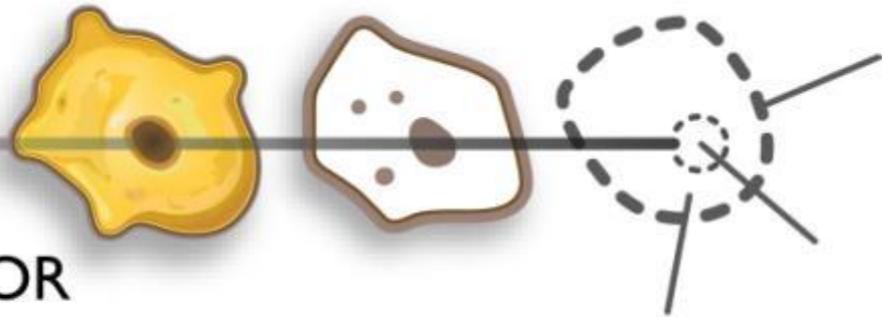


# Using ontology-based annotation to profile disease research

Nigam Shah, MBBS, PhD

[nigam@stanford.edu](mailto:nigam@stanford.edu)



NATIONAL CENTER FOR

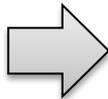
**BIOMEDICAL ONTOLOGY**

# 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.

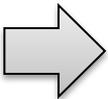
http://rest.bioontology.org

Ontology Services  
Views



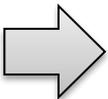
- Download
- Traverse
- Search
- Comment

Mapping Services



- Create
- Download
- Upload

Widgets



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

Annotation



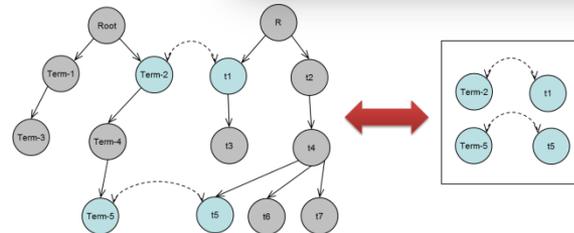
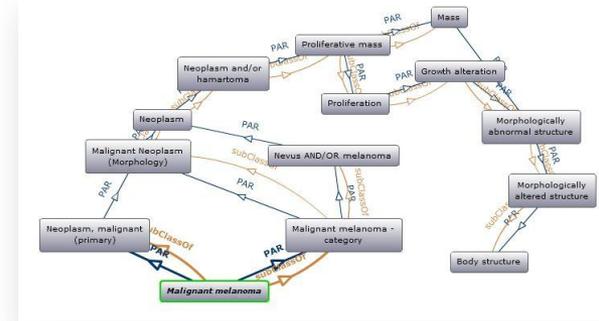
Term recognition

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

Search the Resource Index

To begin, type in text and select a matching term. Examples: melanoma, lupus, breast cancer, ... Click search and select a repository.

Ontology filters

- 1155 ARRS GoldMiner
- 16444 ArrayExpress
- 101606 ClinicalTrials.gov
- 396 Database of Genet
- 23287 Gene Expres
- 21305 Online
- 832
- 1002
- 11507
- 10
- 19461 UniProt KB
- 895

22 sources

3.5M elements

200 ontologies

2.4M terms

4M mappings

16.4 Billion Annotations Indexed

1847 Pathway Commons

1634 PharmGKB [Drug]

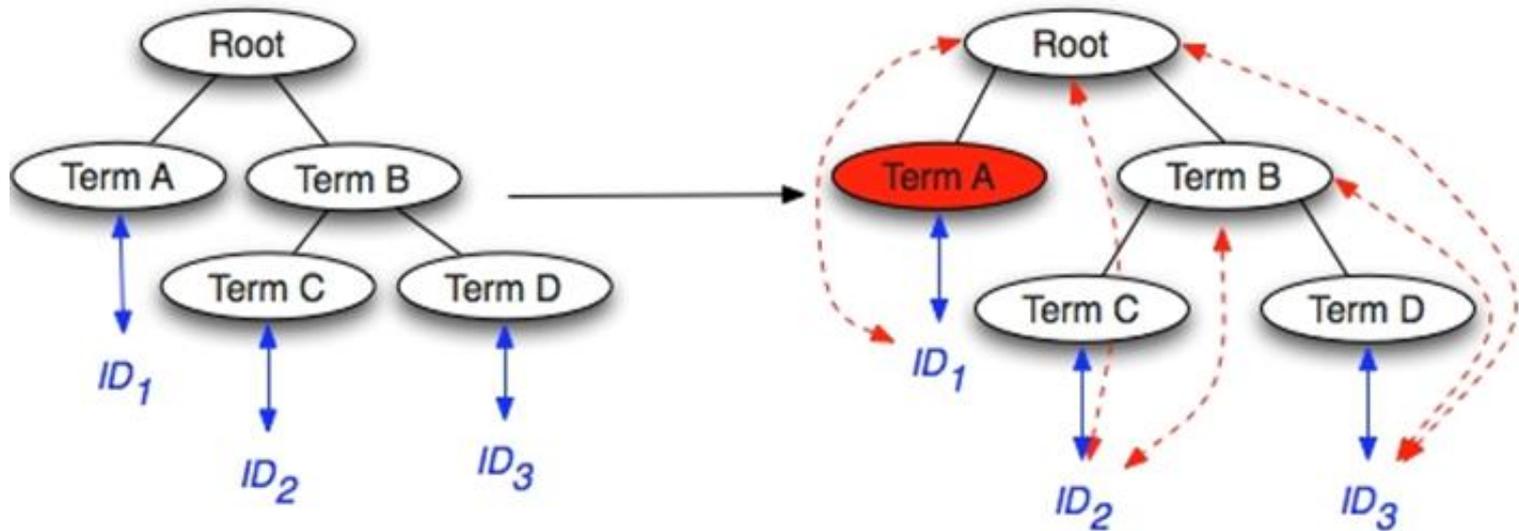
115228 PubChem

127169 Reactome

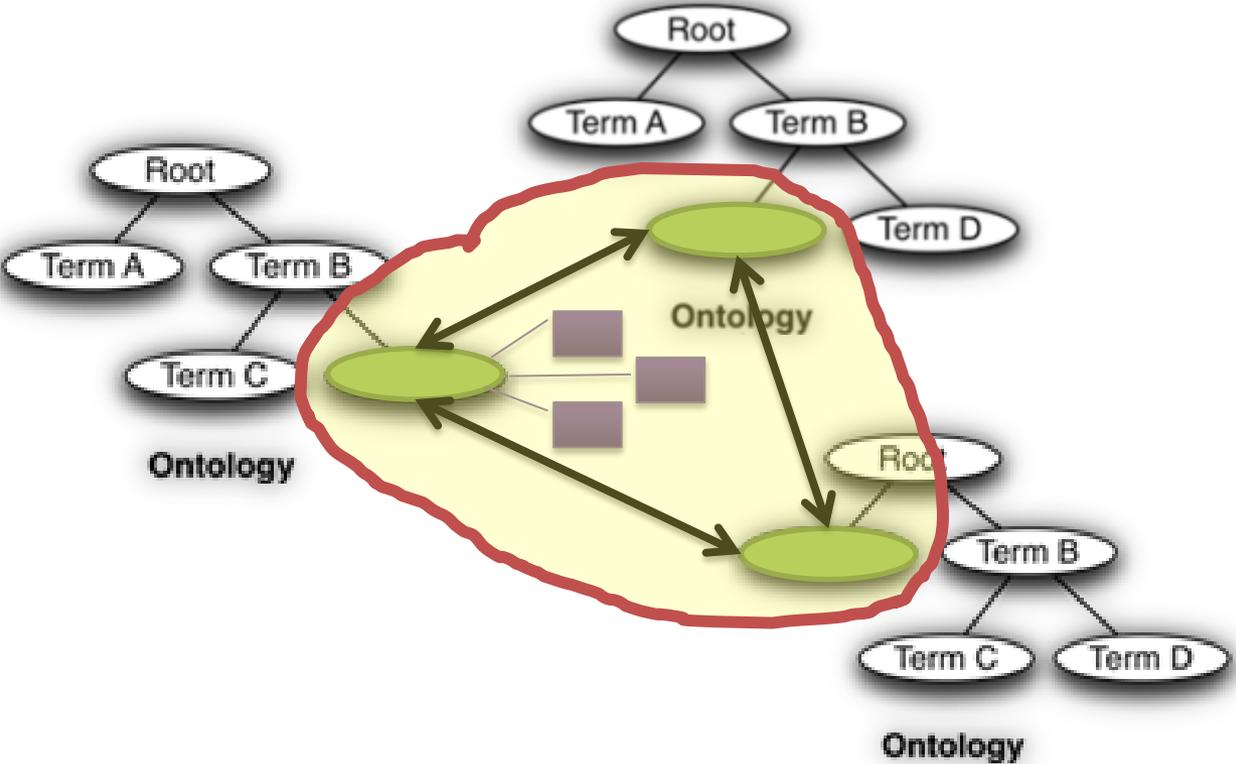
18775 Stanford Microarray Database

1472 WikiPathways

# Transitive closures



# Normalization of alternative IDs



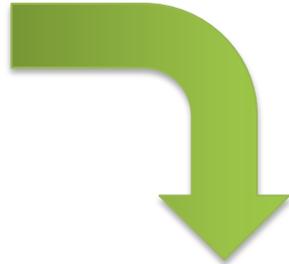
normalized_cui	original_cui	rel	rela	sab
C0000039	C0000039	SY	permuted_term_of	MSH
C0000039	C0216971	SY	same_as	SNOMEDCT



# Annotation service

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

GDS Summary			
Accession:	GDS906 <a href="#">View Expression (GEO profiles)</a>		
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: <a href="#">15467014</a>		
Sample organism:	Homo sapiens	Platform organism:	Homo sapiens
Feature count:	22283	Value type:	count
Series:	<a href="#">GSE1595</a>	Series published:	07/25/2004
Last GDS update:	12/20/2004		

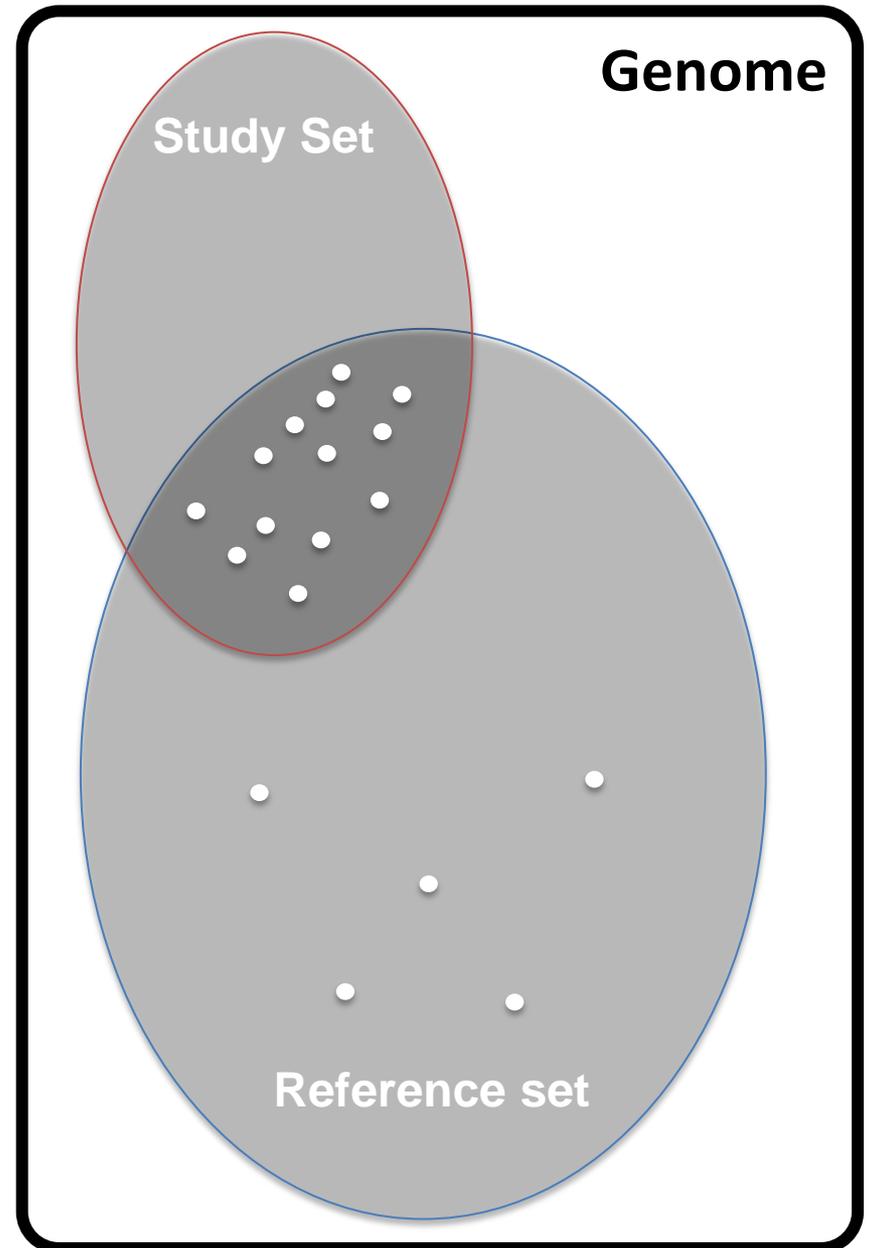


>100 million calls,  
>900 GB of data

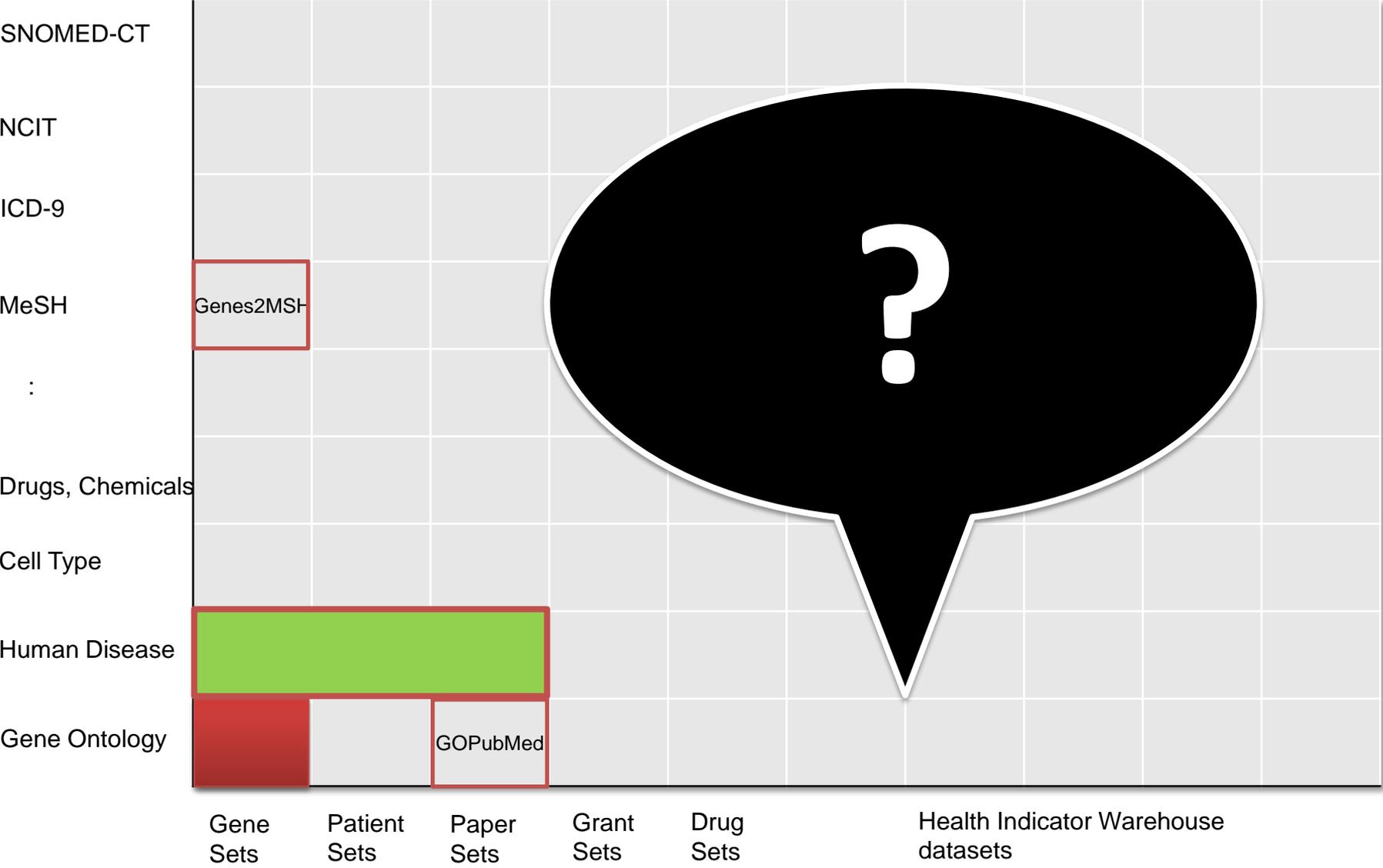
Expression, Expression of bladder, bladder, smooth, bladder muscle, muscle, smooth muscle, cells, mechanical, mechanical stimulation, stimulation, Chronic, results, bladder overdistension, associated, associated with, with, loss, genes, altered

# Generic annotation analysis routine

- Get annotations for each gene in a set
- Count the occurrence of each annotation term in the study set
- Count the occurrence of that term in some reference set (whole genome?)
- P-value for how surprising their overlap is.

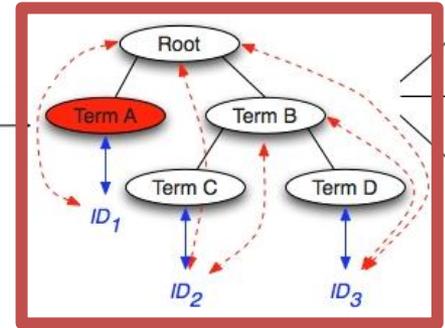
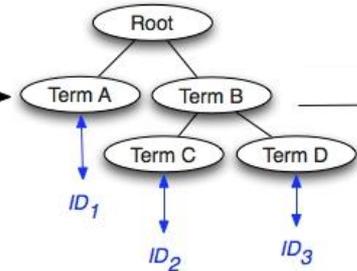
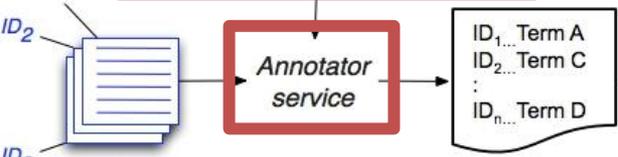
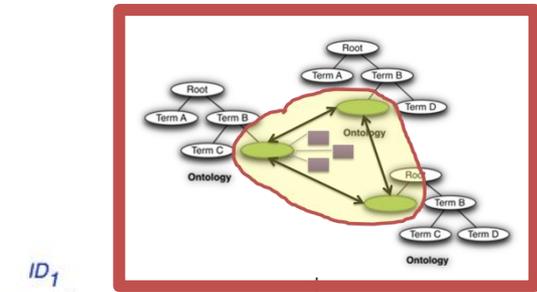


# Annotation Analytics Landscape

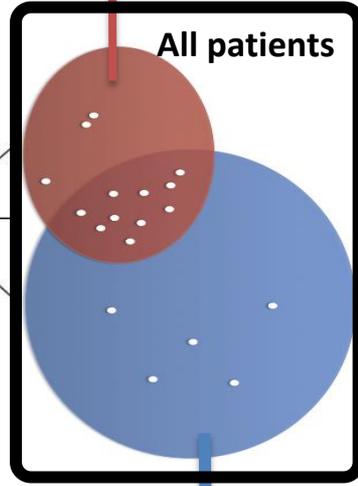




# Profiling a patient set



Patients with abdominal pain



De-ided version of Note\_id: 1058133

PATIENT NAME: X

ADMISSION DIAGNOSIS: Cervical spondylosis at C5-C6 with disk hemiation on the left side and left-sided C6 radiculopathy.

DISCHARGE DIAGNOSIS: Cervical spondylosis at C5-C6 with disk hemiation on the left side and left-sided C6 radiculopathy.

PROCEDURES PERFORMED: C5-C6 anterior cervical disectomy and fusion with instrumentation, performed by Y.

**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 spondylosis at C5-C6 with disk hemiation.**

PAST MEDICAL HISTORY: Hypertension.

PAST SURGICAL HISTORY: None.

MEDICATIONS ON ADMISSION:

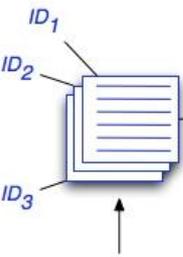
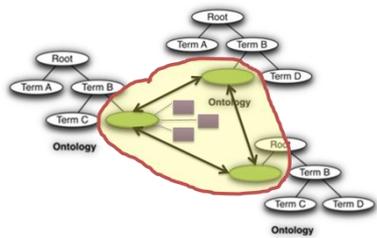
1. Advair Diskus one puff b.i.d.
2. Aciphex 20 mg one tablet q.a.m.
3. Flovent MDI two puffs p.o. b.i.d.
4. Albuterol MDI two puffs q.4-6h. p.r.n.
5. Flonase two sprays per nostril q.d.
6. Nexium 40 mg one capsule q.d.

PHYSICAL EXAMINATION ON ADMISSION: GENERAL: The patient had clear consciousness and no acute distress. NEUROLOGIC: The patient's range of motion was 70% of normal, but there was no Lhermitte sign and no Spurling sign. There were no focal weaknesses present for both upper and lower extremities. Brachioradialis was 1/4 on both sides. Triceps was 1/4 to 2/4 on both sides. Knee jerk and ankle jerk were 2/4 on both sides. Posterior tibialis was 2/4 on both sides. LUNGS: Clear to auscultation bilaterally. CARDIOVASCULAR: Regular rate and rhythm. ABDOMEN: Benign. EXTREMITIES: No cyanosis or edema.

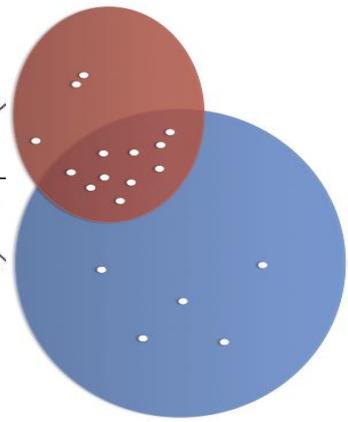
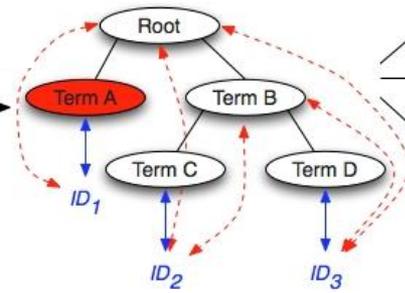
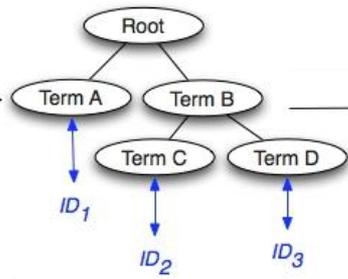
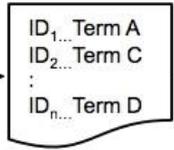
HOSPITAL COURSE: The patient was admitted to the hospital on 2003, and taken to the operating room for a C5-C6 ACDF with instrumentation. The patient tolerated the procedure well. There were no complications. The patient recovered well in the PACU without any events, and the patient was transferred to the floor. While on the floor there were no acute events overnight. The patient was tolerating oral intake, and the patient's pain was under control with oral pain medications. On postoperative day #1 the patient was tolerating an oral diet, ambulating and voiding on his own, and the patient was discharged home.



Appropriate control

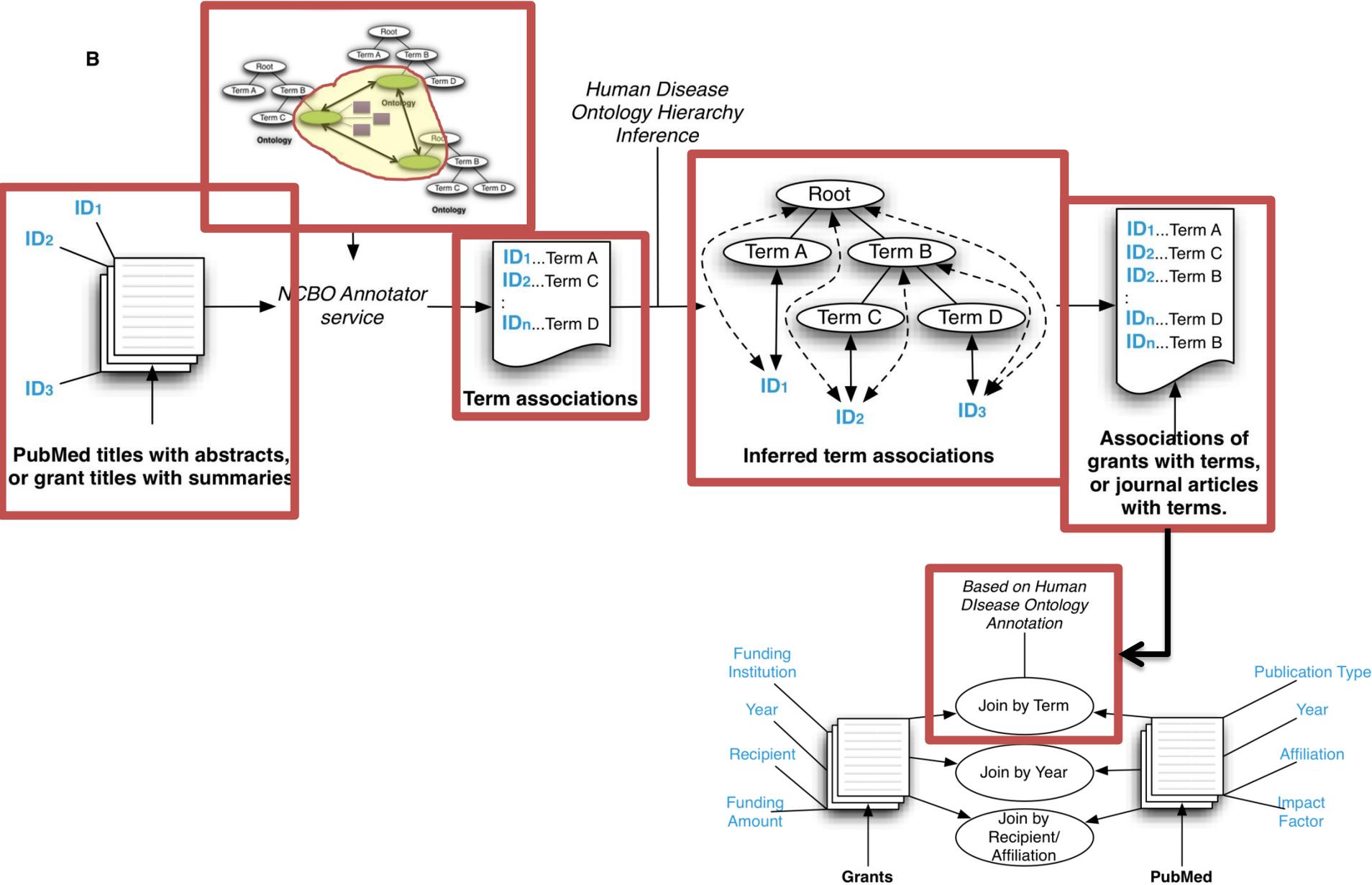


Annotator service



Grants

# Profiling Annotations of Grants, Publications



# Data

## Grants from 1997 to 2007

33 funding agencies

## Publications from Medline

Only “Journal articles”

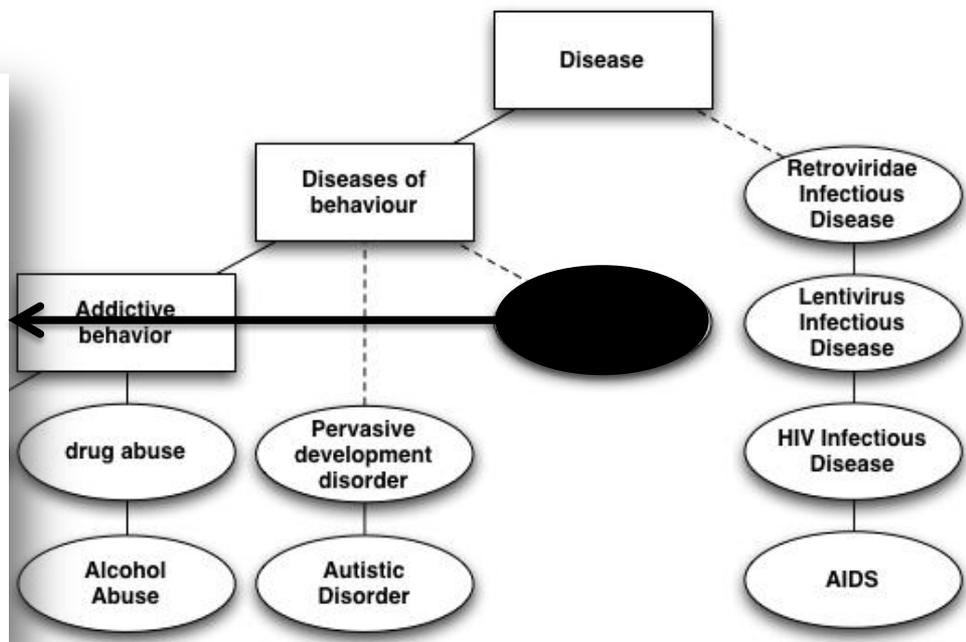
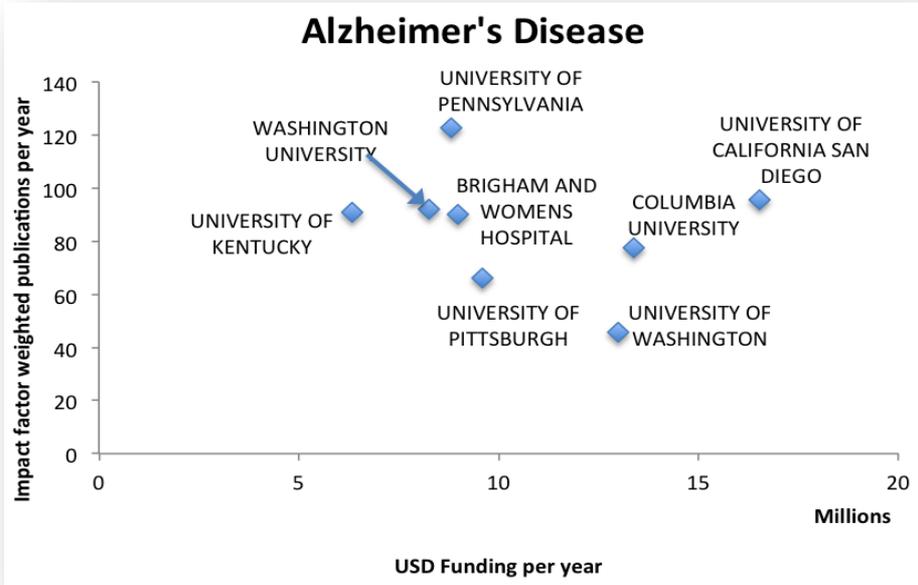
Funding Institution	Acronym
Agency for Healthcare Research and Quality	AHRQ
Center for Disease Control and Prevention	CDC
Congressional Liaison Committee of the Coalition for Life Sciences	CLC
Department of Defense	DOD
US Food and Drug Administration	FDA
Fogarty International Center	FIC
Health Resource and Service Administration	HRSA
National Aeronautics and Space Administration	NASA
National Center for Complementary and Alternative Medicine	NCCAM
National Cancer Institute	NCI
National Institute on Minority Health and Health Disparities	NCMHD
National Center for Research Resources	NCRR
National Eye Institute	NEI
National Human Genome Research Institute	NHGRI
National Heart, Lung, and Blood Institute	NHLBI
National Institute on Aging	NIA
National Institute on Alcohol Abuse and Alcoholism	NIAAA
National Institute of Allergy and Infectious Diseases	NIAID
National Institute of Arthritis and Musculoskeletal and Skin Diseases	NIAMS
National Institute of Biomedical Imaging and Bioengineering	NIBIB
Eunice Kennedy Shiver National Institute of Child Health and Human Development	NICHD
National Institute on Drug Abuse	NIDA
National Institute on Deafness and Other Communication Disorders	NIDCD
National Institute of Dental and Craniofacial Research	NIDCR
National Institute of Diabetes and Digestive and Kidney Diseases	NIDDK
National Institute of Environmental Health Sciences	NIEHS
National Institute of General Medical Sciences	NIGMS
National Institute of Mental Health	NIMH
National Institute of Neurological Disorders and Stroke	NINDS
National Institute of Nursing Research	NINR
National Library of Medicine	NLM
National Science Foundation	NSF
The Substance Abuse and Mental Health Services Administration	SAMHSA

- 327 billion USD across 81,858 grants.
- 137 billion USD for grants that are annotated with at least one disease term.
- 2.4 million journal articles.

# Results

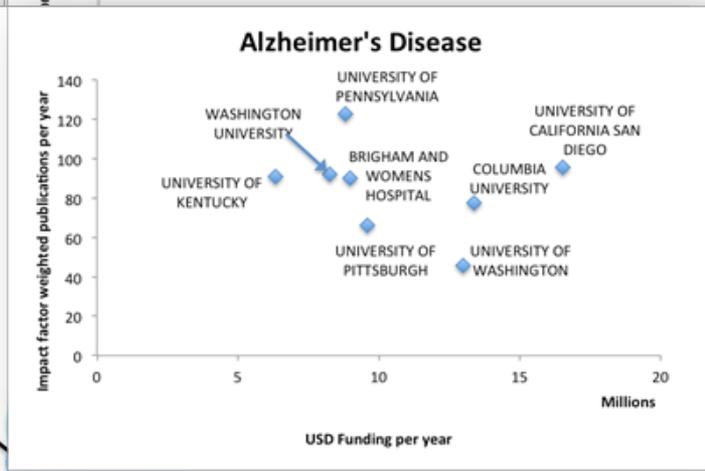
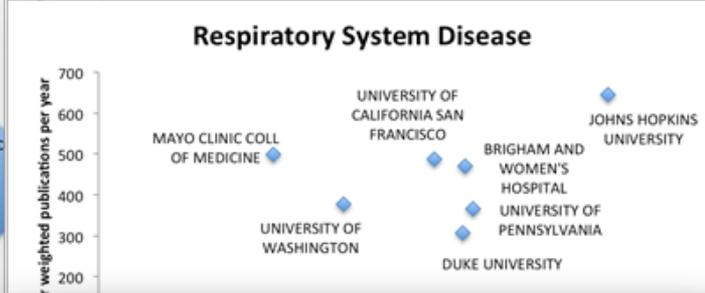
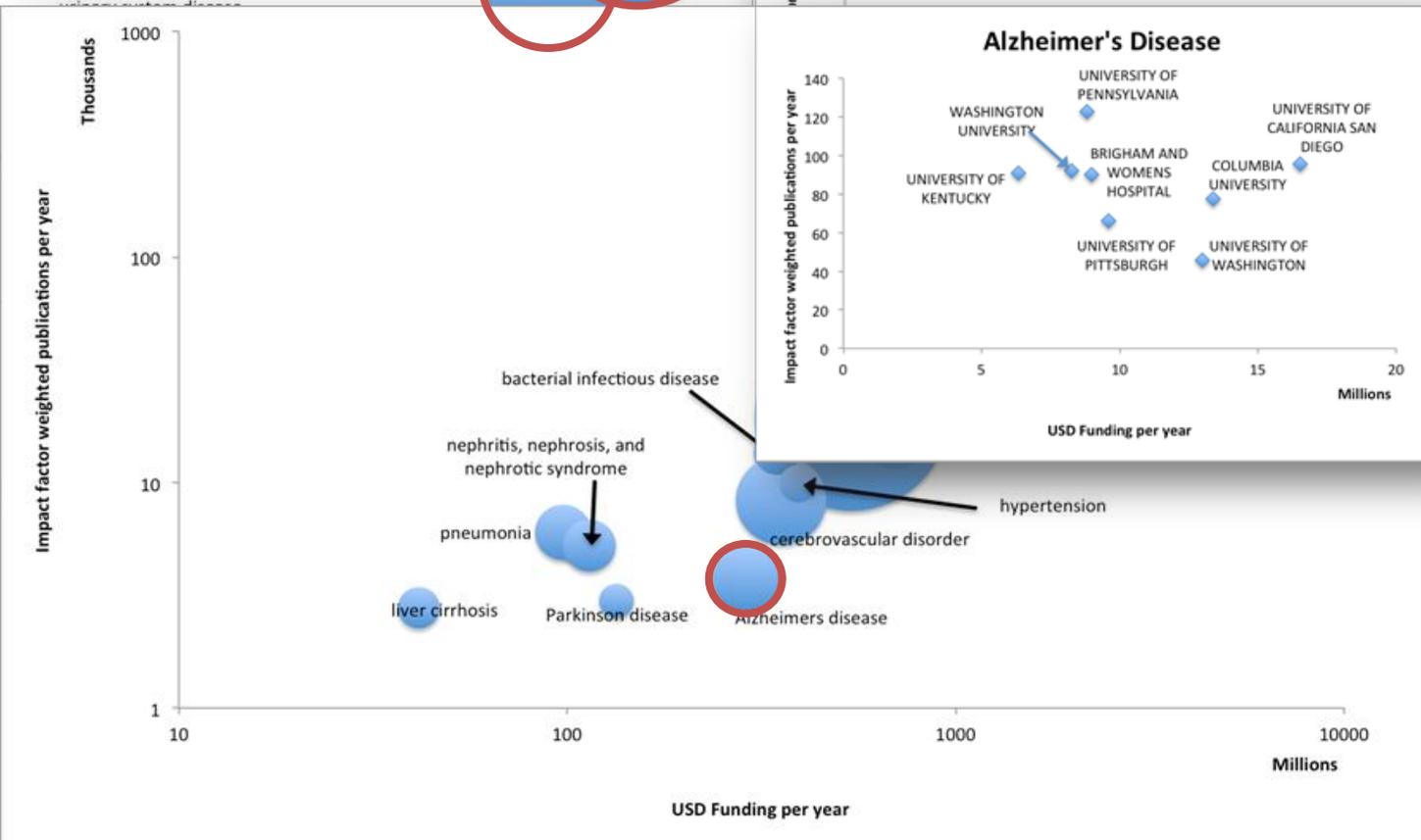
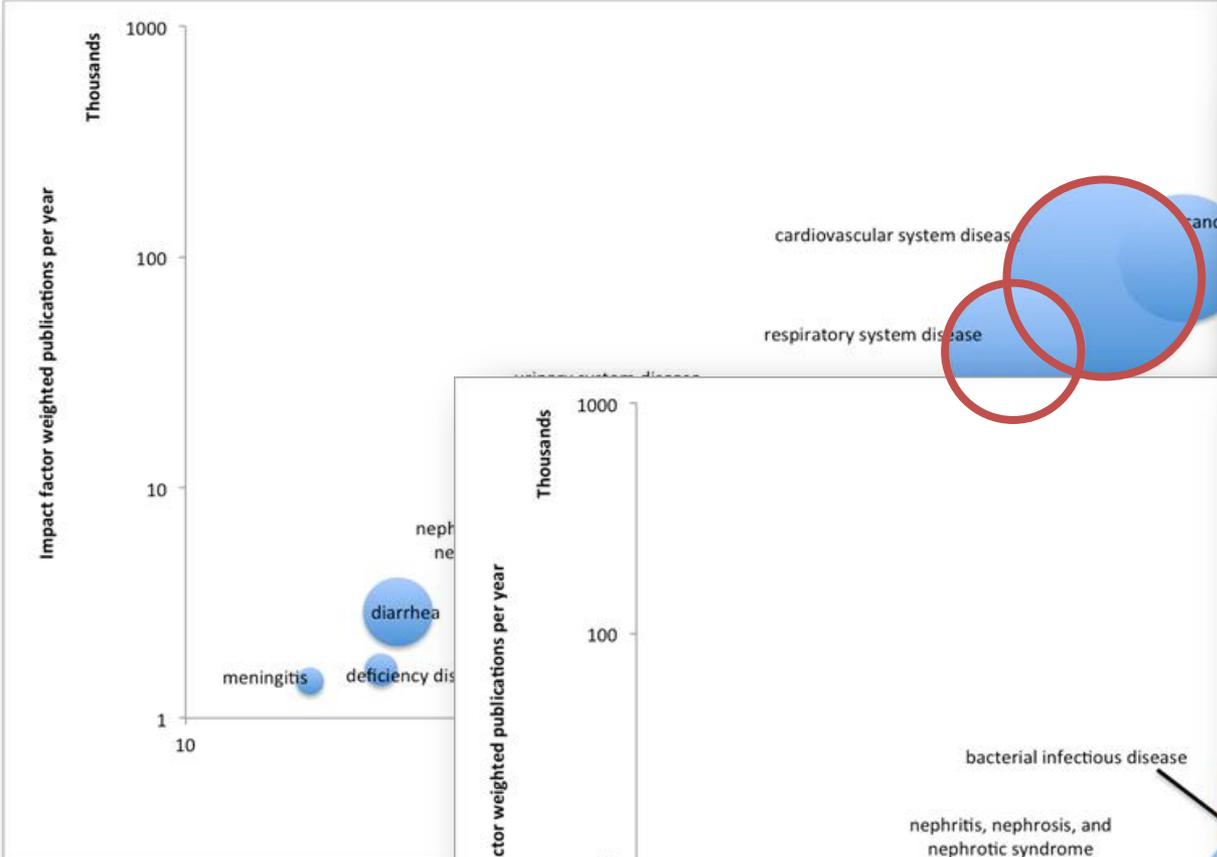
- Sponsorship: USD funding amount divided by the impact factor weighted publication count for a topic.
- Allocation: Is funding proportional to the size of the problem?
- Trends: For a topic over time, across funding institutions

# Sponsorship



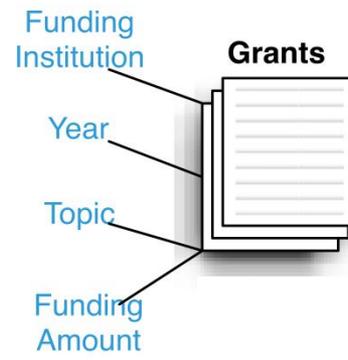
Funding per Year (Millions of USD)	Impact factor-Weighted Publications Per Year (Thousands)	Sponsorship (Thousands of USD per Impact factor-weighted Publication)	DO term
\$338.39	1.33	254.60	drug abuse
\$71.22	0.34	212.22	drug dependence
\$927.90	4.46	208.01	acquired immunodeficiency syndrome
\$1,110.48	8.52	130.27	HIV infectious disease
\$1,113.93	8.60	129.53	Lentivirus infectious disease
\$77.09	0.60	128.84	autistic disorder
\$1,116.19	8.74	127.64	Retroviridae infectious disease
\$66.62	0.56	119.72	alcohol abuse
\$80.63	0.69	116.09	pervasive development disorder
\$289.00	3.76	76.82	Alzheimers disease

# Allocation

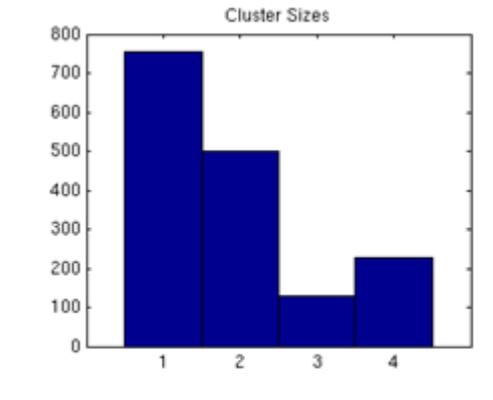
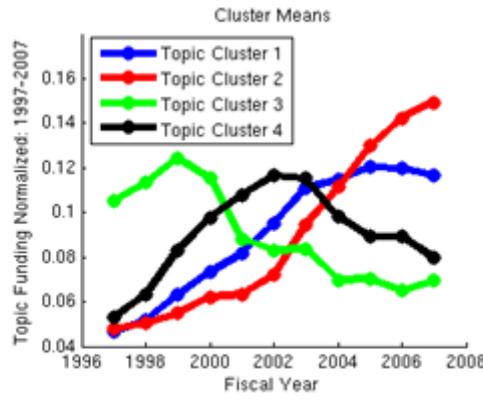
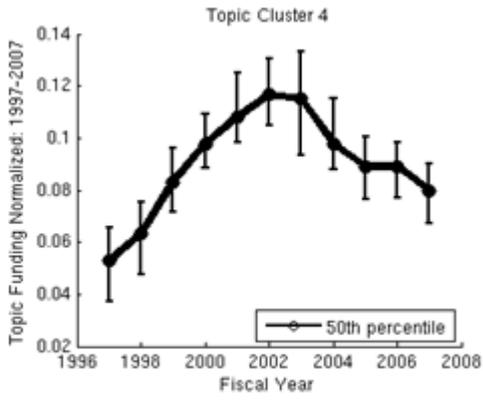
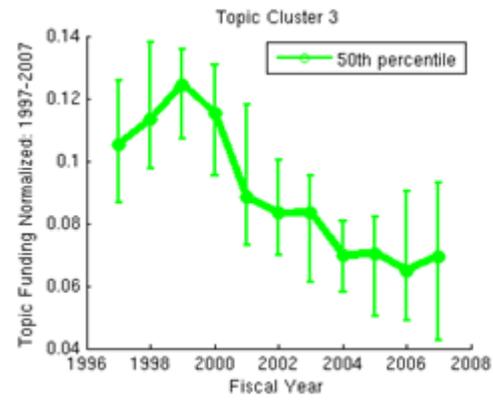
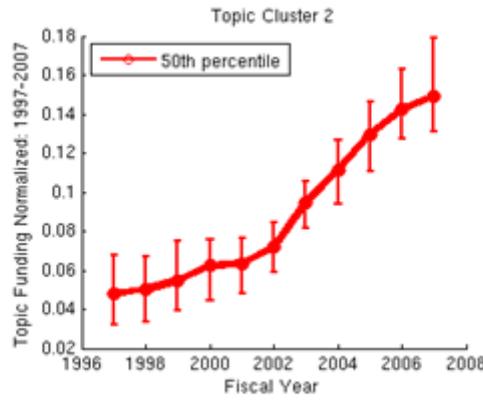
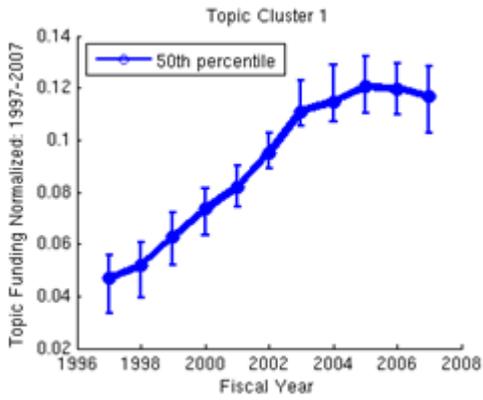


# Trends:

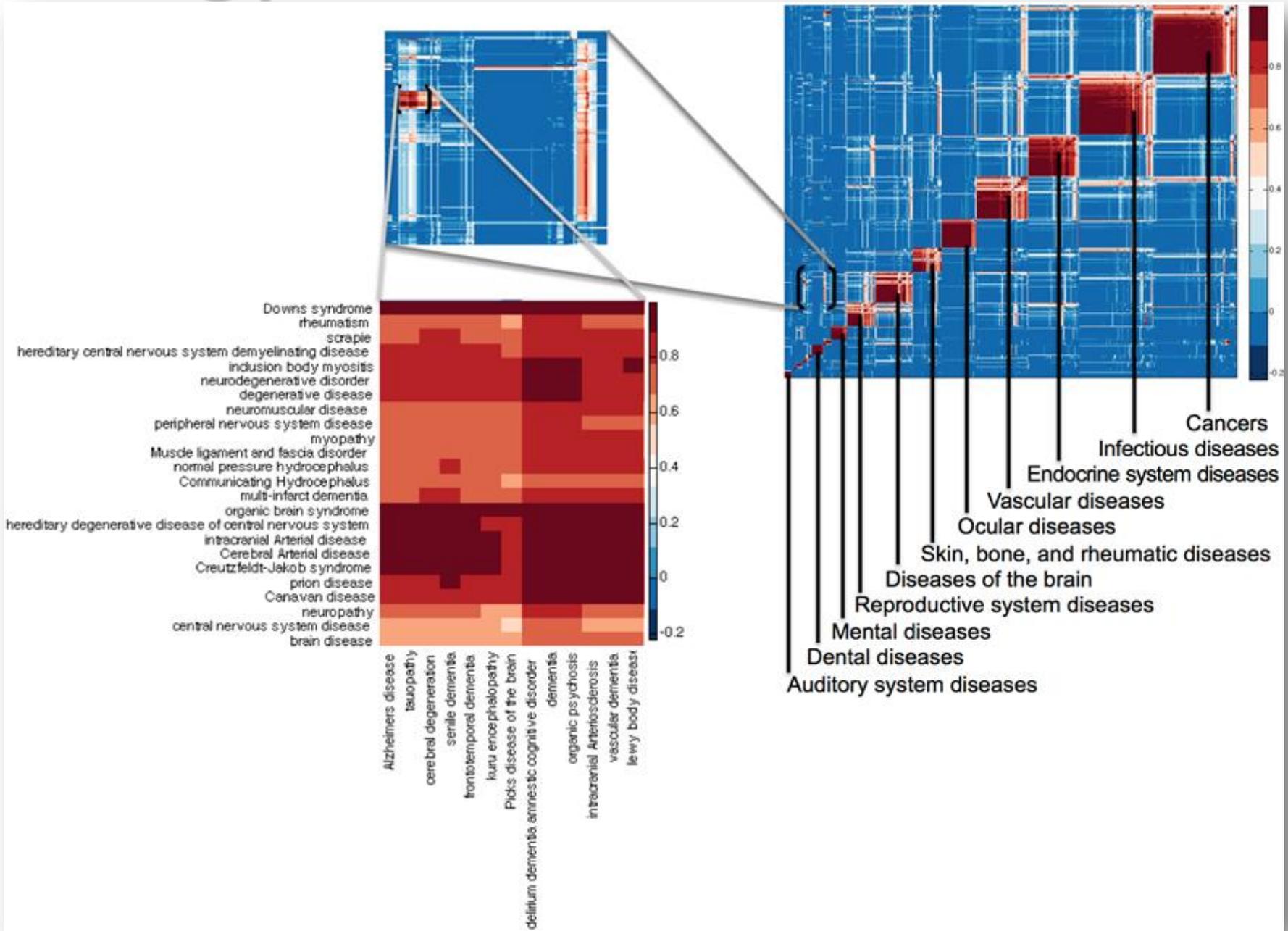
## Who funds what



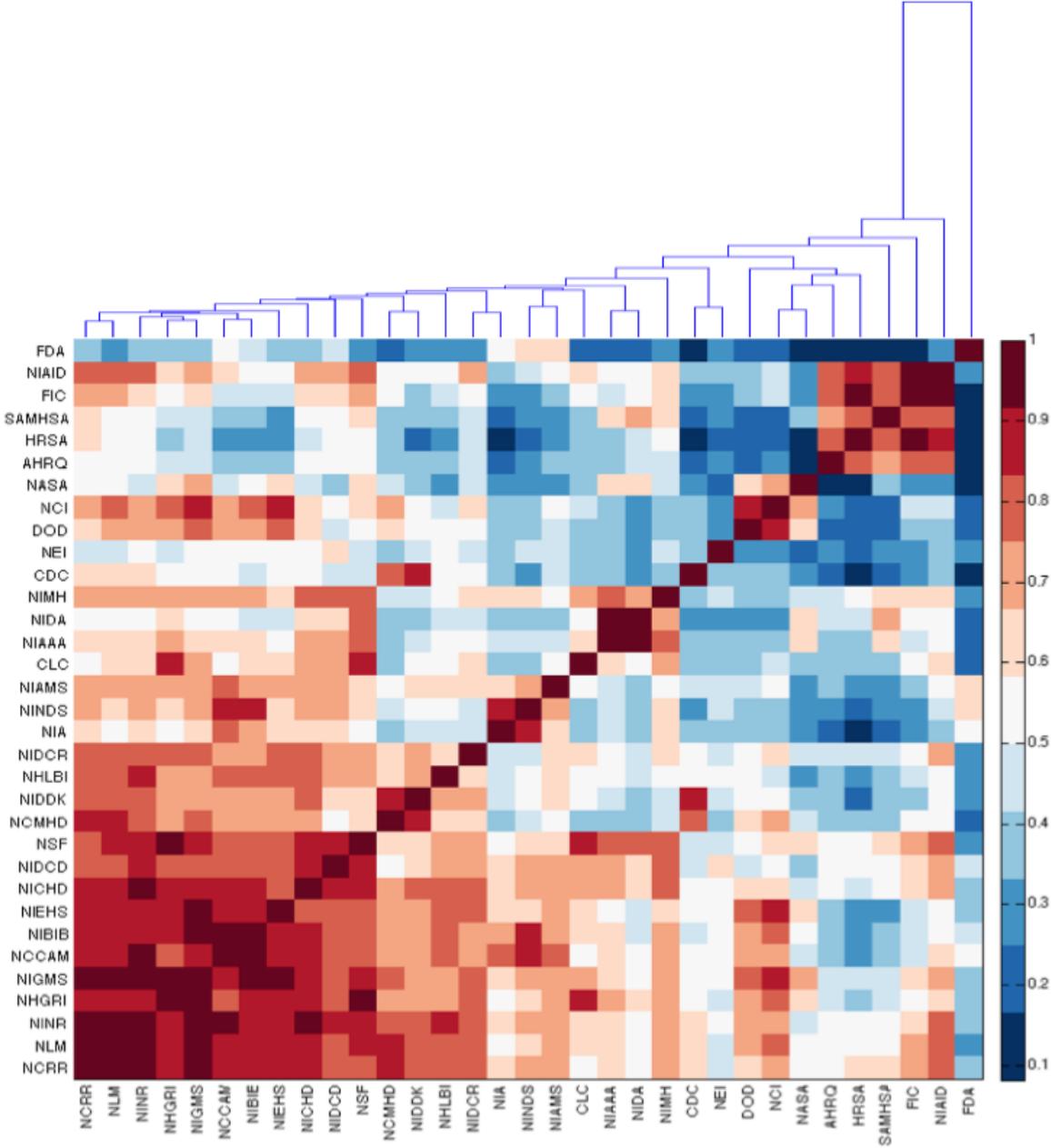
# Trends



# Funding patterns



# Funding patterns



# Sanity checks

# Summary

- Profiling using ontologies is widely used in “omics” research.
- It is possible to annotate grants and publications using automated methods.
- Analysis of the annotations can reveal patterns that help understand the landscape of disease research.

