Using ontology-based annotation to profile disease research

Nigam Shah, MBBS, PhD
nigam@stanford.edu
**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.**
Ontology Services
- Download
- Traverse
- Search
- Comment

Views

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://rest.bioontology.org

http://bioportal.bioontology.org
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 |
Creating Lexicons

Sentence in Medline abstract – 1

Sentence in Medline abstract – n

Term – 1

Term – n

Frequency

Syntactic types

Semantic types

IC filter

SYN filter

STY filter
Annotation service

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

>100 million calls, 
>900 GB of data
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

SNOMED-CT
NCIT
ICD-9
MeSH
::
Drugs, Chemicals
Cell Type
Human Disease
Gene Ontology

Gene Sets  Patient Sets  Paper Sets  Grant Sets  Drug Sets  Health Indicator Warehouse datasets

Genes2MSH
GOPubMed
Profiling a set of Aging genes

Disease Ontology

261 Age-related genes

Genome

~ 30% of genome
Profiling a patient set

Patients with abdominal pain

Appropriate control

Patients with abdominal pain

All patients
Grants
Profiling Annotations of Grants, Publications

- PubMed titles with abstracts, or grant titles with summaries
- NCBO Annotator service
- Term associations
- Human Disease Ontology Hierarchy Inference
- Based on Human Disease Ontology Annotation
- Join by Term
- Join by Year
- Join by Recipient/Affiliation
- Funding Institution
- Year
- Recipient
- Funding Amount
- Impact Factor
- Publication Type
- Affiliation
- Grants
- PubMed

Inferred term associations
Associations of grants with terms, or journal articles with terms.
### Grants from 1997 to 2007

33 funding agencies

| 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 | NCIMHD |
| 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 Shriver 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 |

### Publications from Medline

Only “Journal articles”

- 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

Alzheimer's Disease

- Impact factor weighted publications per year
- USD Funding per year

Disease
- Diseases of behaviour
- Addictive behavior
- Drug abuse
- Pervasive development disorder
- Alcohol Abuse
- Autistic Disorder

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 | Lentiviridae 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
• 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.
• Paea LePendu
• Yi Liu
• Adrien Coulet

• NIH – NCBO funding
• Scholarly Database
  • Katy Borner
• Research Crossroads
  • Kyle Brown
• Alex Skrenchuk