NLM and Visual Analytics: A partnership for discovery and engagement

Patricia Flatley Brennan, RN, PhD
Director
National Library of Medicine

NLM Strategic Plan
Transforming Data into Knowledge

Accelerate discovery and advance health through data-driven research

Reach more people in more ways through enhanced dissemination and engagement

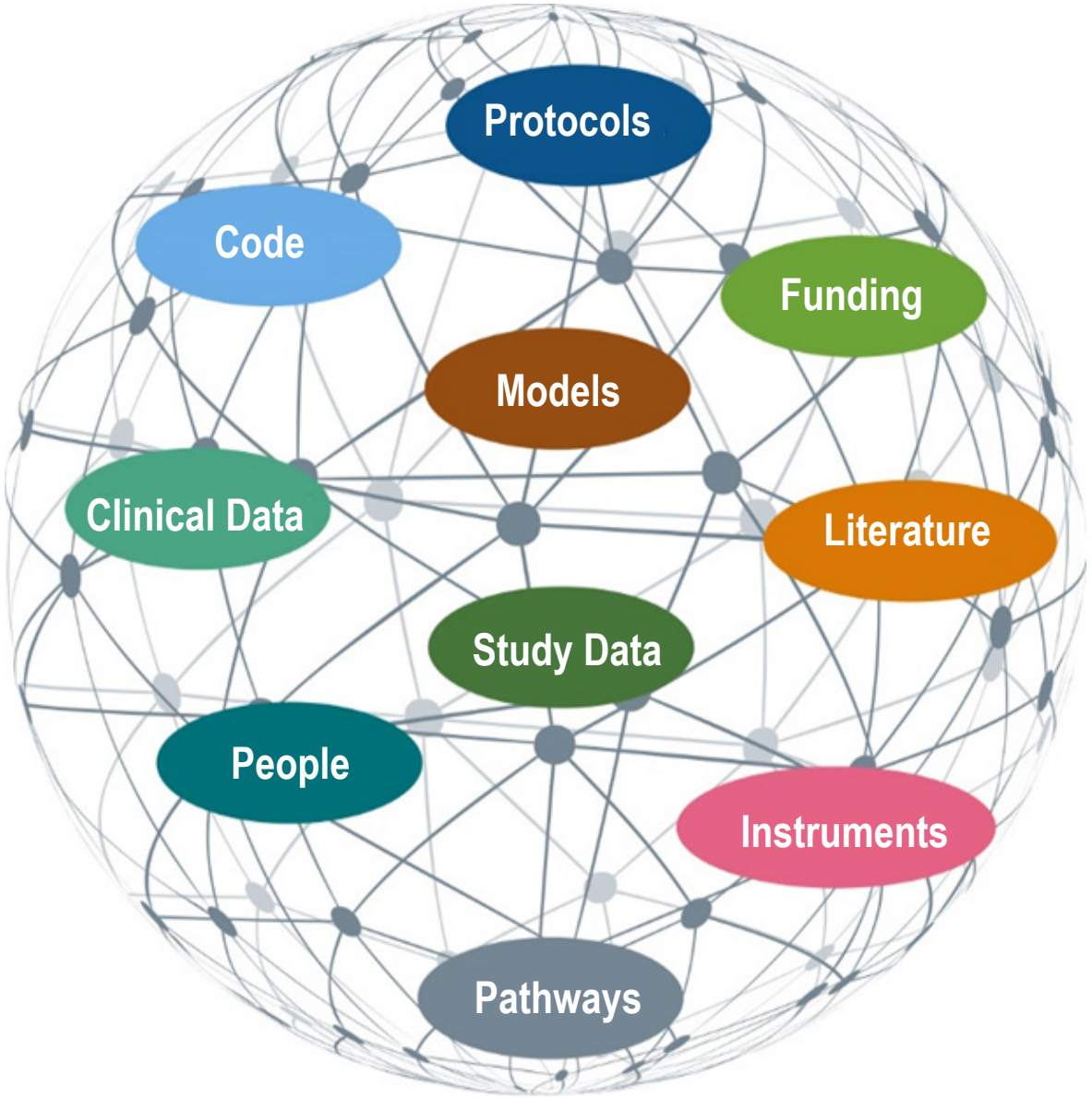
Build a workforce for data-driven research and health

Accelerate discovery and advance health through data-driven research

NIH NLM

Fostering a ecosphere of discovery

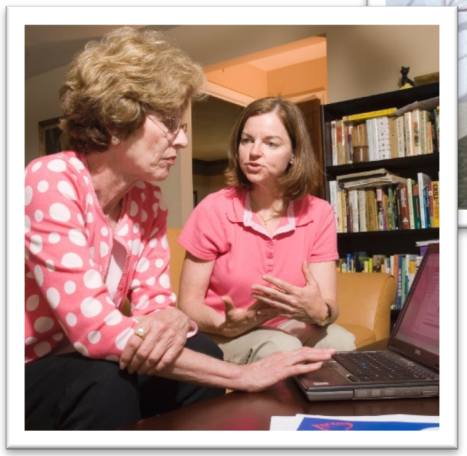
digital research objects

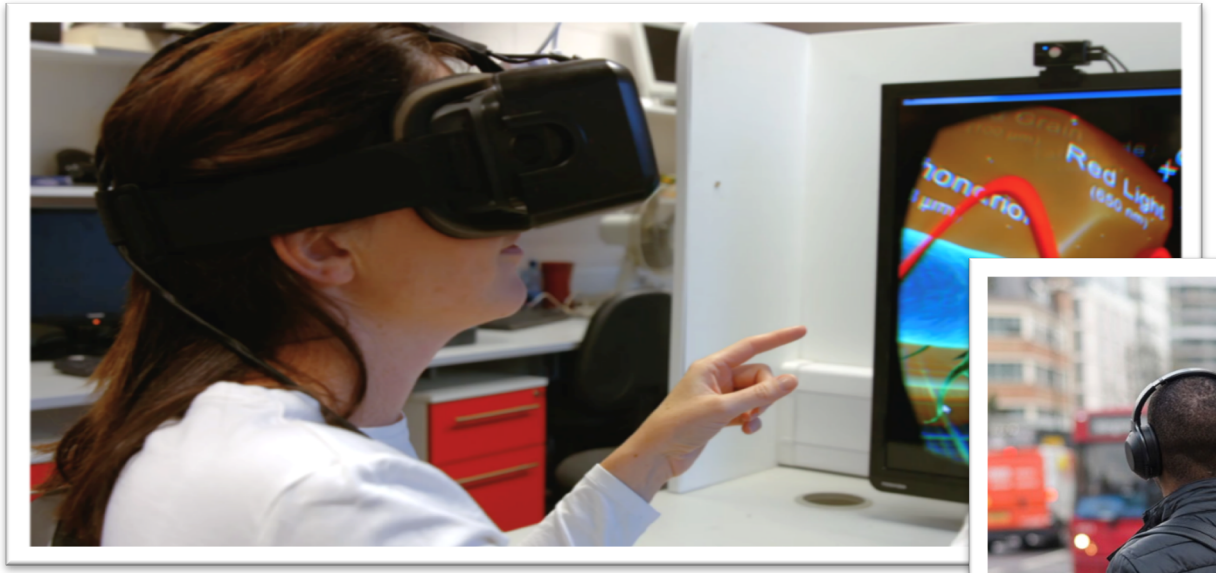


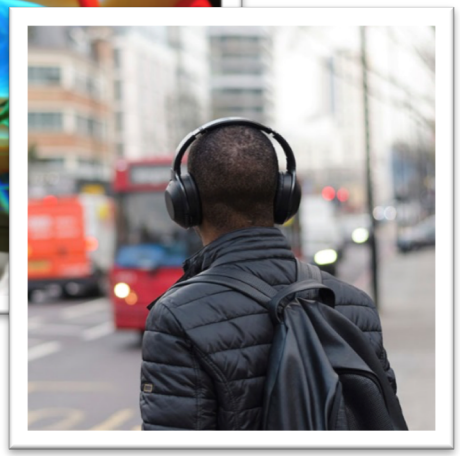
Reach more people in more ways through enhanced dissemination and engagement

New users, New ways

Biomedical & health information access methods & information dissemination strategies







Foster distinctiveness of NLM as a

reliable, trustable source

of health information & biomedical data

Build a workforce for data-driven research and health

Expand & enhance research training

for biomedical informatics & data science

The Human Enterprise: Reaching People Wherever

NLM & Visual Analytics: A Partnership for Discovery and Engagement

- Accelerate use of NLM resources
- Stimulate discovery in VAHC
- Promote sustainability of VAHC strategies, utilities and tools

Accelerate use of NLM resources

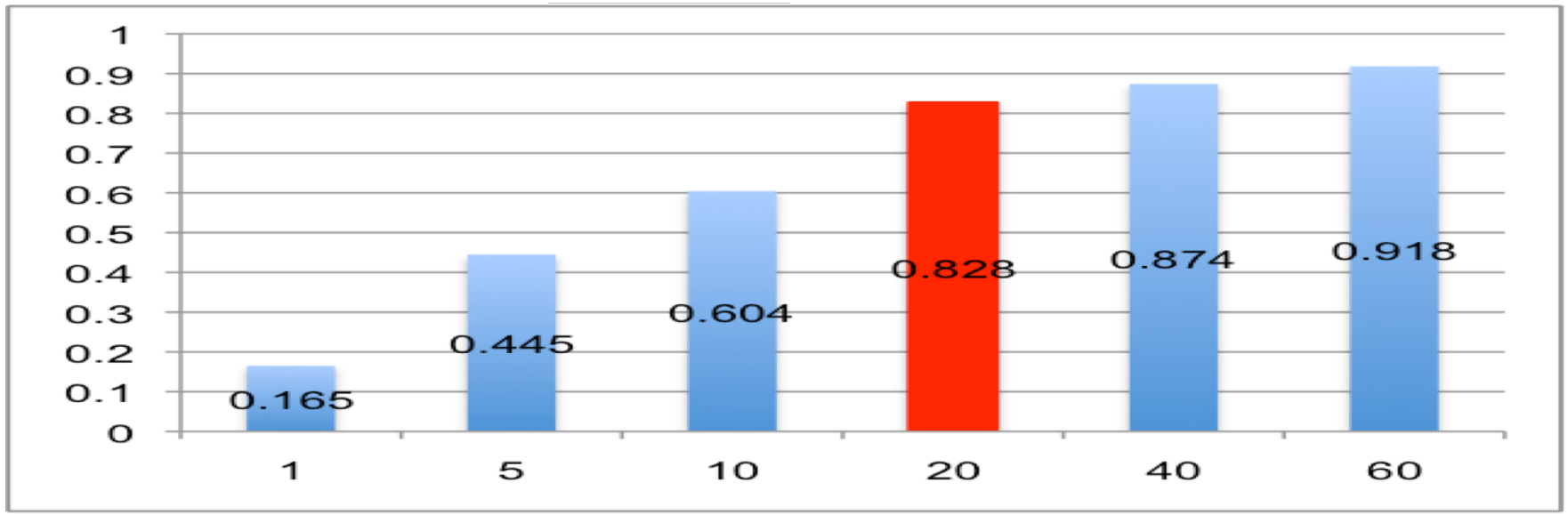
The Literature Challenge

Image: goo.gl/FLCjZP

Today's user search behavior

"The best place to hide a dead body is page 2 of search results"

Google



Most (>80%) clicks happened in top 20 positions.

Over half of PubMed queries return more than 20 results.

PubMed labs

Search

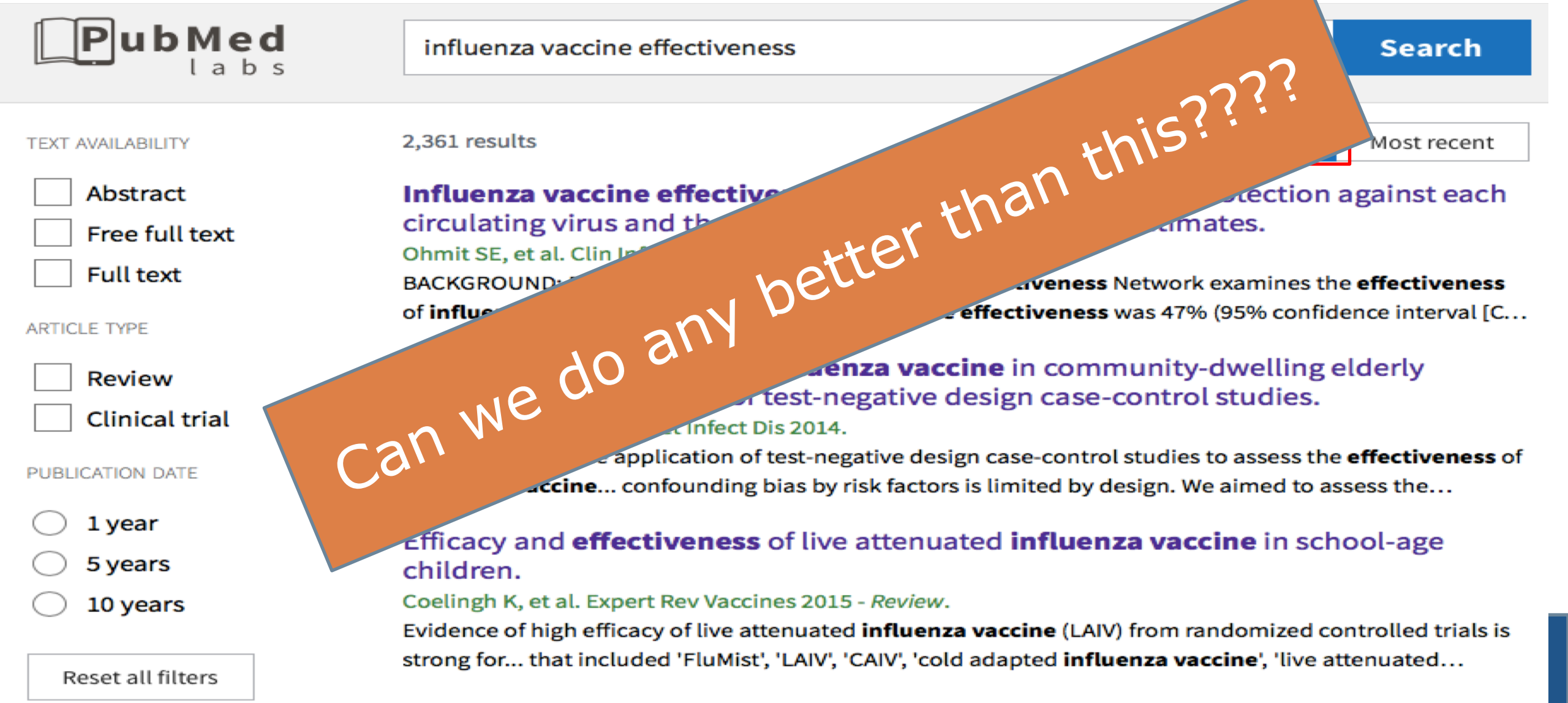
Try otitis media treatment koonin e crispr influenza vaccine effectiveness

What is PubMed Labs?

PubMed Labs is a test site where we are *experimenting* with new features and tools that eventually may be incorporated in PubMed, in their current or a revised form based on the input we receive. Please try the site and let us know what you think.

Feedback

Search Results Page



Can we do any better than this????

Daily Data Flow at NCBI

The Hubble Telescope generates 10TB of data … *per year*

10-15 Terabytes (TB)/day

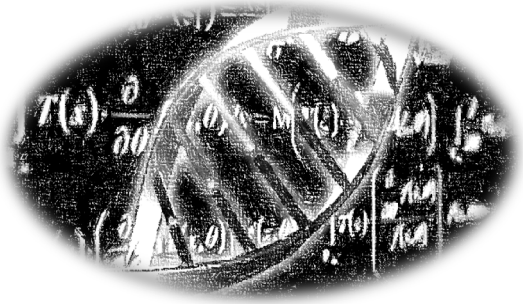
Data Submissions

NCBI Internal Processing

550 Terabytes/day

200 Petabytes/year

20 TB/day



Interactive Web

95 TB/day

Data Downloads

1 Petabyte = 1,000 TB = 1,000,000 GB = 1,500,000 CD-ROMs

GenBank was last released on CD-ROM in 1996

Stimulate discovery in VAHC

RFI Next-Generation Data Science Challenges

(NOT LM 17-006)

Data-Driven Health Improvement

1. Make EHRs better at the point of care
    1. **Reduce fragmentation in patient health records and simplify visualization for care givers**
    2. Personally-tailored decision systems to help patients
    3. Automate integration of personal data from mobile devices into clinical workflows
2. Extract research value of electronic health records
    1. Algorithms based on patient similarity to drive CER
    2. **Real time structuring of unstructure**d text in electronic health records
    3. Mobile health data monitoring systems that protect PHI
    4. Address **bias in health records** that are used for research purposes
    5. Nuanced phenotypes including severity, degree and certainty

Advanced Data Analytics Management

- Improved ontologies, vocabularies and standards
  - Improved ontologies and crosswalks between information capture systems
  - Common data models and standard vocabularies
  - Create crosswalks among Common Data elements/models
  - Expand personal health data beyond health record to population-level health determinants
- Better, more accessible methods that extend to the cloud
  - Open source analytic and simulation methods
  - Leverage AI and semantic analytics to integrate data sources across cloud services
  - **Technology platforms** to support data storage and analysis by scientists
  - Statistical tools that are more programming oriented than descriptive packages but easy to learn and deploy quickly
  - Develop accurate, privacy-producing linking methods
  - Study how technology has affected **documentation** of data analysis and information methods
- Curation at scale
  - Automated assignment of standardized metadata to existing datasets and digital files
  - Automatic assignment of metadata early in the research process
  - **Automate integration** of personal data from mobile devices into clinical workflows

Intelligent Learning Systems

- Approaches for **engaging users with their health data**
- Learner-centric modalities with **content 'chunking'** to modularize declarative knowledge and procedural experiences
- Brain science research focused on **learning and retention**

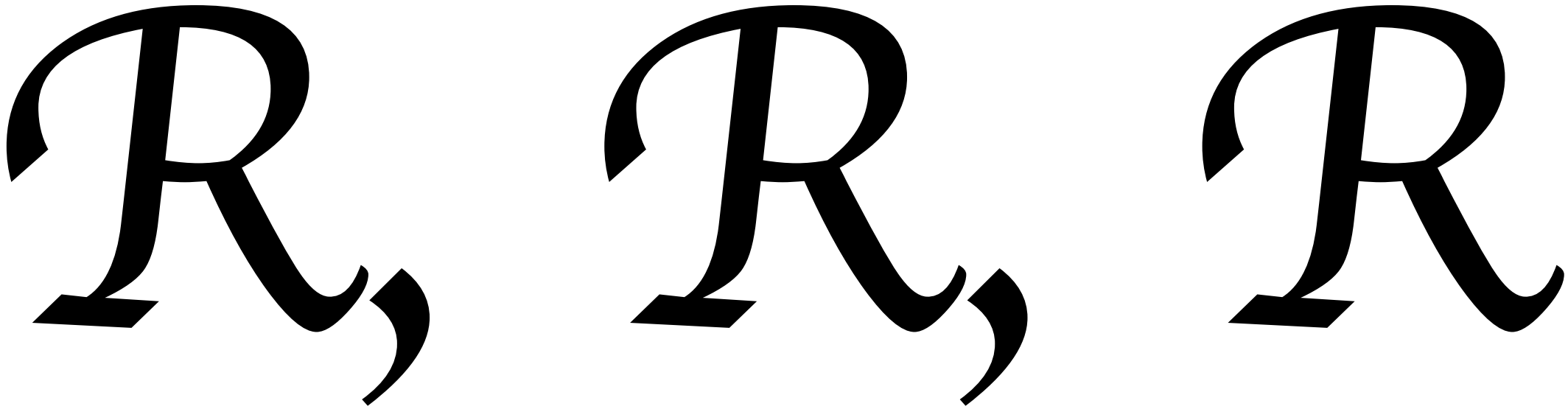
Promote sustainability of VAHC strategies, utilities and tools

DATA

DISCOVERY | BEST PRACTICES | LINKS TO LIT | COLLECTION | PROMOTION | IMPACT

Do for data what we've done for the literature.

Why does anyone care about models?



Rigor, Reproducibility, Reuse

Curator and Custodian: The NLM Collection

What does a library of VAHC tools look like?

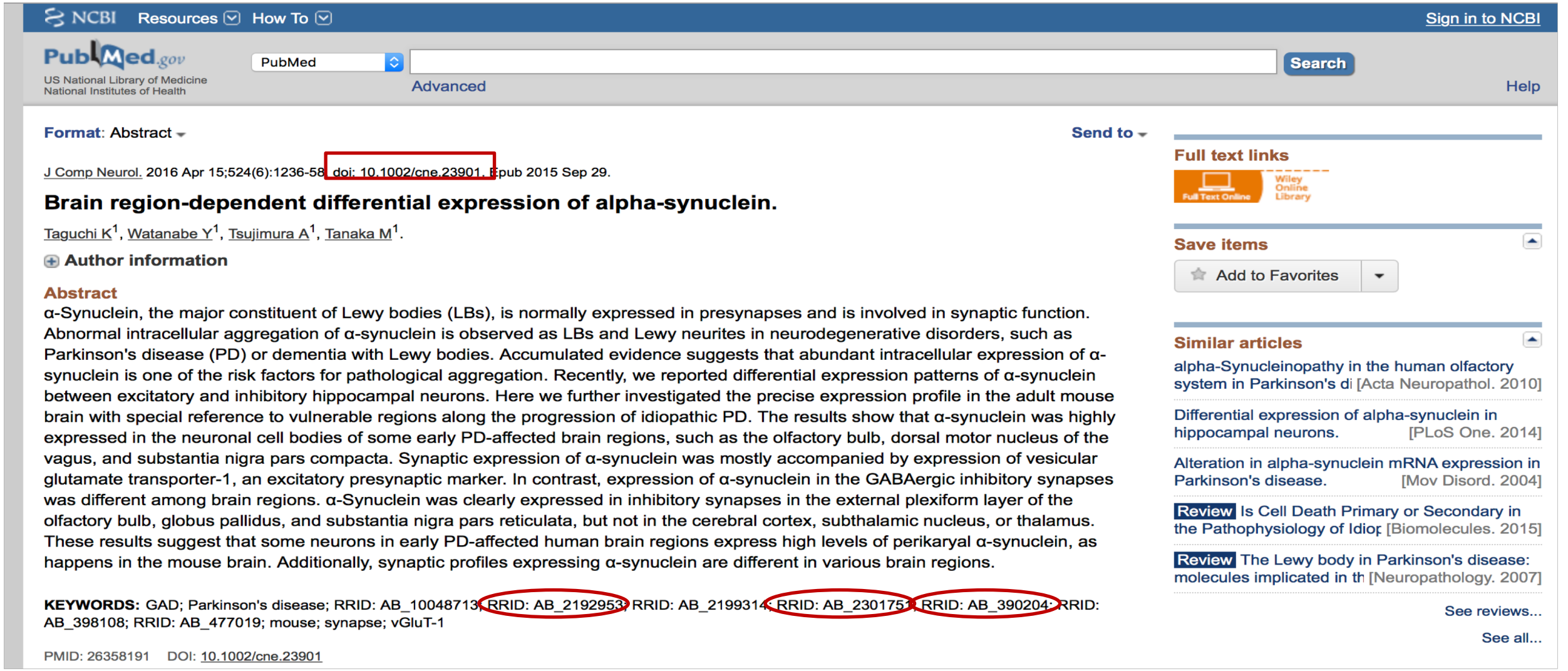
RRIDs characterize Methods

- identify the model organisms, cells lines, antibodies, and tools (such as software or databases) you have used
- Blogs, software took kits, algorithm implementation
- include Research Resource Identifiers (RRIDs) within the materials and methods section of their papers

doi's characterize objects

- Complete declarations (objects, articles, books)
- ISO standard, computational registry

RRIDs & doi's in PubMed



NCBI Resources How To Sign in to NCBI

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed Search
Advanced Help

Format: Abstract Send to

J Comp Neurol. 2016 Apr 15;524(6):1236-58. doi: 10.1002/cne.23901. Epub 2015 Sep 29.

Brain region-dependent differential expression of alpha-synuclein.

Taguchi K[1], Watanabe Y[1], Tsujimura A[1], Tanaka M[1].

Author information

Abstract

α-Synuclein, the major constituent of Lewy bodies (LBs), is normally expressed in presynapses and is involved in synaptic function. Abnormal intracellular aggregation of α-synuclein is observed as LBs and Lewy neurites in neurodegenerative disorders, such as Parkinson's disease (PD) or dementia with Lewy bodies. Accumulated evidence suggests that abundant intracellular expression of α-synuclein is one of the risk factors for pathological aggregation. Recently, we reported differential expression patterns of α-synuclein between excitatory and inhibitory hippocampal neurons. Here we further investigated the precise expression profile in the adult mouse brain with special reference to vulnerable regions along the progression of idiopathic PD. The results show that α-synuclein was highly expressed in the neuronal cell bodies of some early PD-affected brain regions, such as the olfactory bulb, dorsal motor nucleus of the vagus, and substantia nigra pars compacta. Synaptic expression of α-synuclein was mostly accompanied by expression of vesicular glutamate transporter-1, an excitatory presynaptic marker. In contrast, expression of α-synuclein in the GABAergic inhibitory synapses was different among brain regions. α-Synuclein was clearly expressed in inhibitory synapses in the external plexiform layer of the olfactory bulb, globus pallidus, and substantia nigra pars reticulata, but not in the cerebral cortex, subthalamic nucleus, or thalamus. These results suggest that some neurons in early PD-affected human brain regions express high levels of perikaryal α-synuclein, as happens in the mouse brain. Additionally, synaptic profiles expressing α-synuclein are different in various brain regions.

KEYWORDS: GAD; Parkinson's disease; RRID: AB_10048713; RRID: AB_2192953; RRID: AB_2199314; RRID: AB_2301751; RRID: AB_390204; RRID: AB_398108; RRID: AB_477019; mouse; synapse; vGluT-1

PMID: 26358191 DOI: 10.1002/cne.23901

Full text links
Full Text Online Wiley Online Library

Save items
Add to Favorites

Similar articles
alpha-Synucleinopathy in the human olfactory system in Parkinson's di [Acta Neuropathol. 2010]
Differential expression of alpha-synuclein in hippocampal neurons. [PLoS One. 2014]
Alteration in alpha-synuclein mRNA expression in Parkinson's disease. [Mov Disord. 2004]
Review Is Cell Death Primary or Secondary in the Pathophysiology of Idiop [Biomolecules. 2015]
Review The Lewy body in Parkinson's disease: molecules implicated in th [Neuropathology. 2007]

See reviews...
See all...

Questions for discussion

1. What does a library of visual analytic tools look like?
2. Who would use it?
3. How does it differ from models, protocols, pathways, etc?
4. How is authority imbued?
5. Should models be stored separately from the data they are used on and the reports they are used for?
6. How can the NLM help advance VAHC?

Reaching NLM

@NLM_news
@NLMdirector

patti.brennan@nih.gov