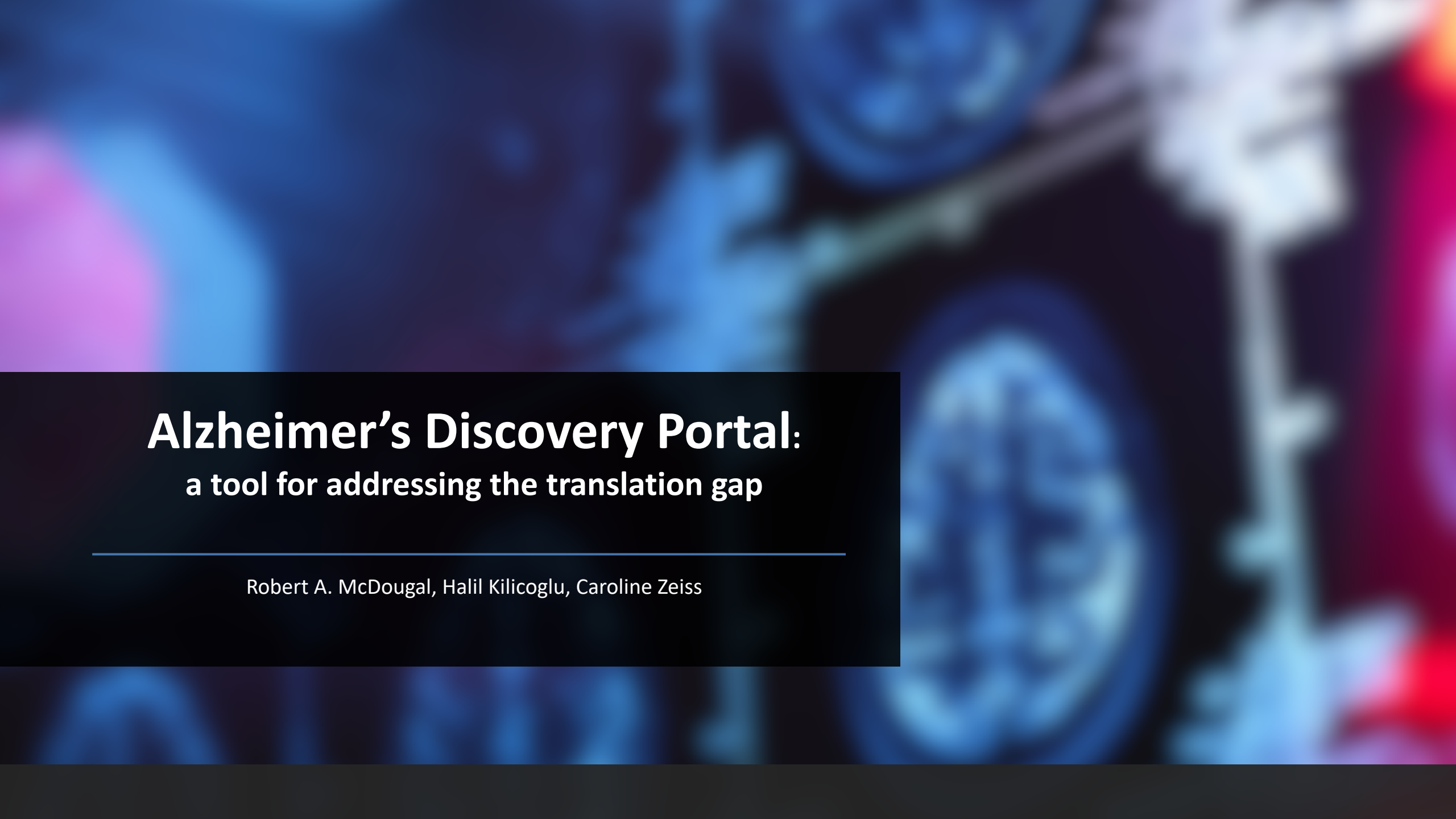


Alzheimer's Discovery Portal:

a tool for addressing the translation gap

Robert A. McDougal, Halil Kilicoglu, Caroline Zeiss

Another major drug candidate targeting the brain plaques of Alzheimer's disease has failed. What's left?

By **Kelly Servick** | Mar. 21, 2019 , 6:00 PM

Couzin-Frankel J. When mice mislead. Science. 2013

The majority of interventional studies in animals have positive outcomes, but fail to translate to human trials

Approximately 90% of the drugs reported to have clinical efficacy in high profile journals based on animal models fail in clinical trials

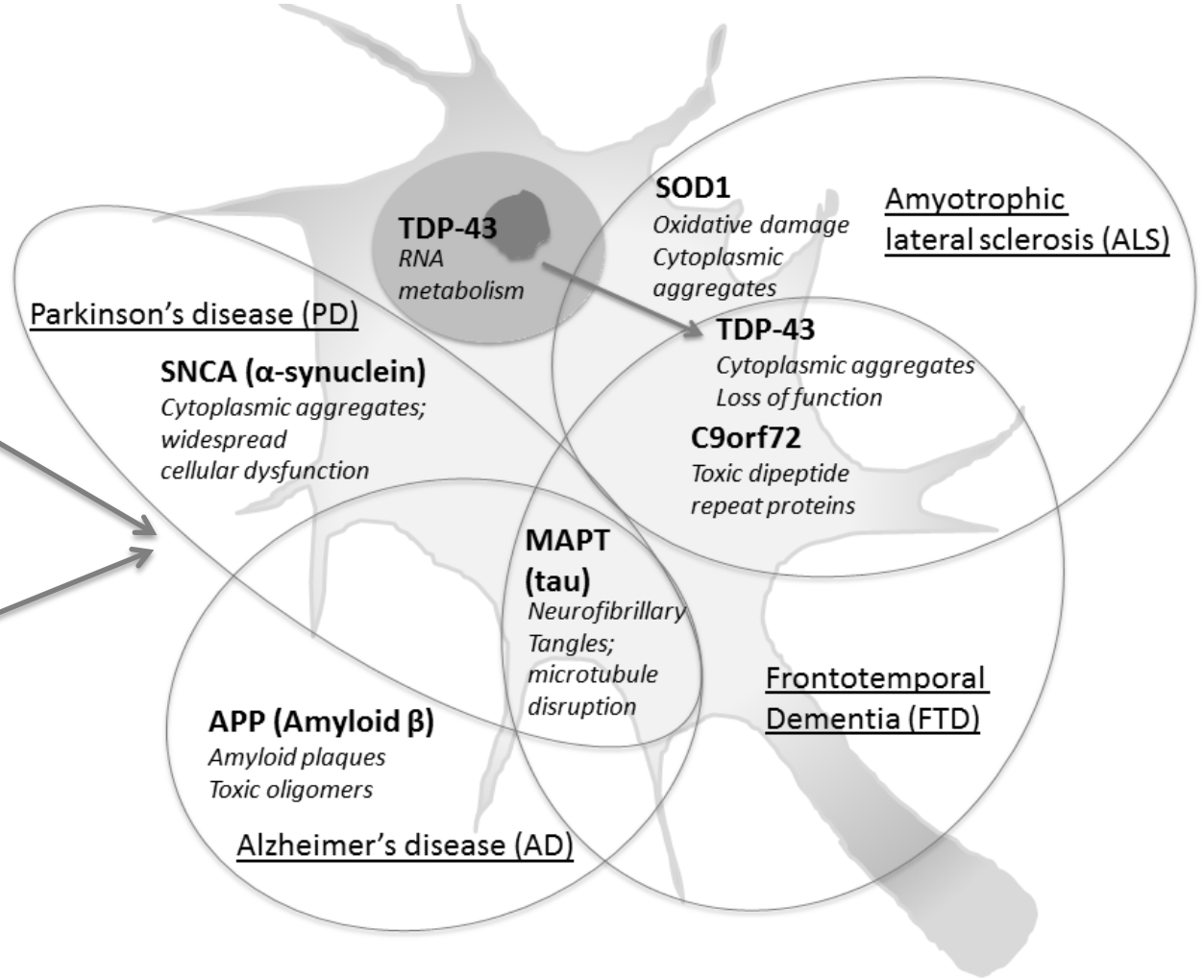
Translatability vs Reproducibility

Using Reductionist Models to Examine Complex Systems

Autosomal Dominant AD
APP, PSEN1, PSEN2

Late Onset AD
Many susceptibility genes

CC or DO mice, or
natural dementias

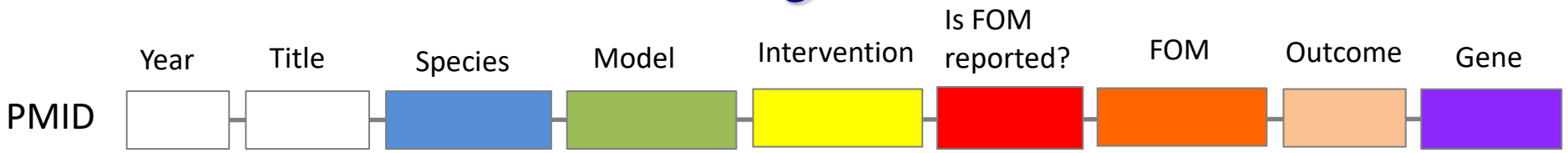


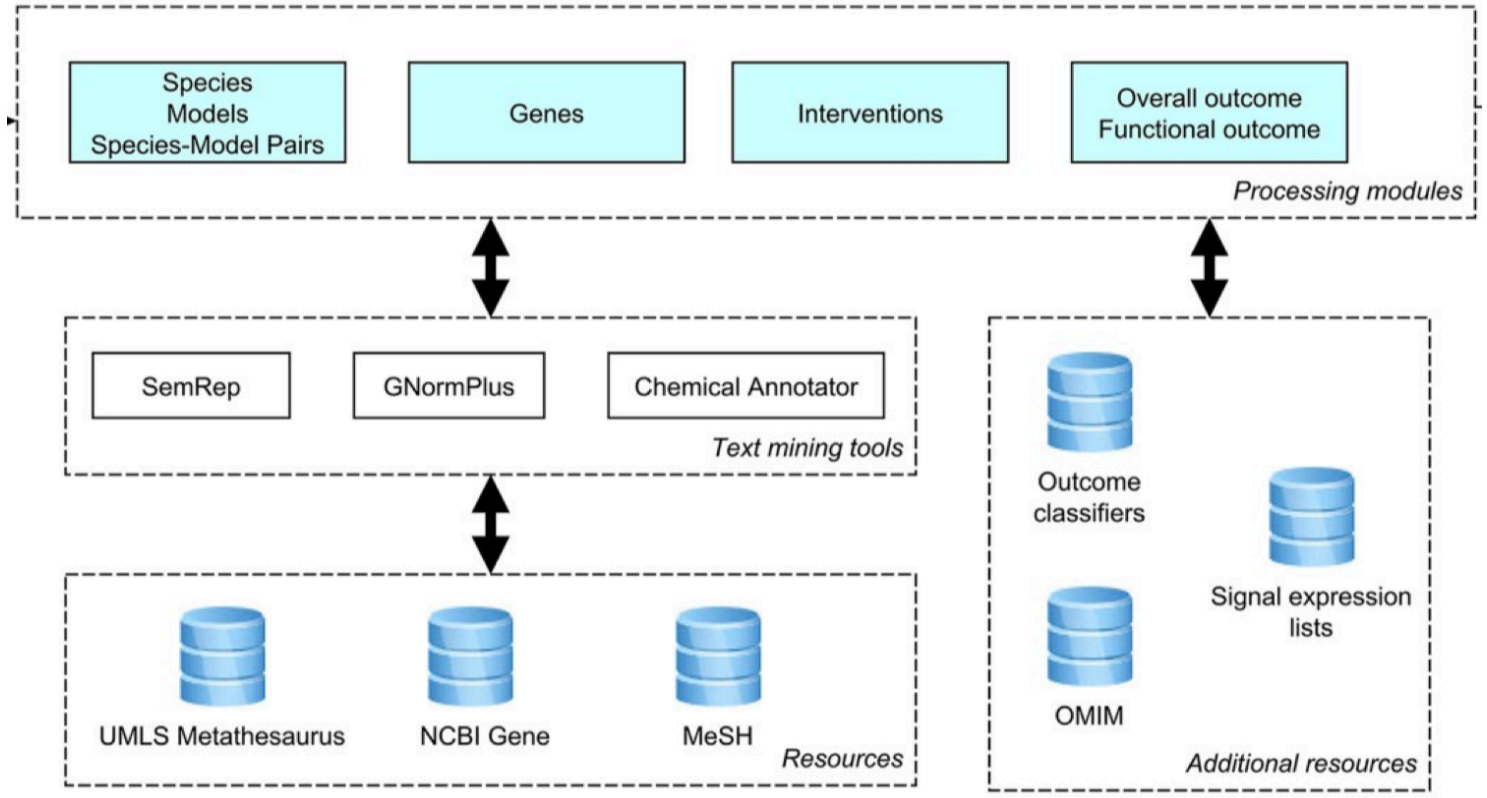
Zeiss CJ. From Reproducibility to Translation in Neurodegenerative Disease. ILAR J. 2017

ALZHEIMERS: Zeiss CJ. Improving the predictive value of interventional animal models data. Drug Discov Today. 2015

PARKINSONS: Zeiss CJ, Allore HG, Beck AP. Established patterns of animal study design undermine translation of disease-modifying therapies for Parkinson's disease. PLoS One. 2017

Menagerie





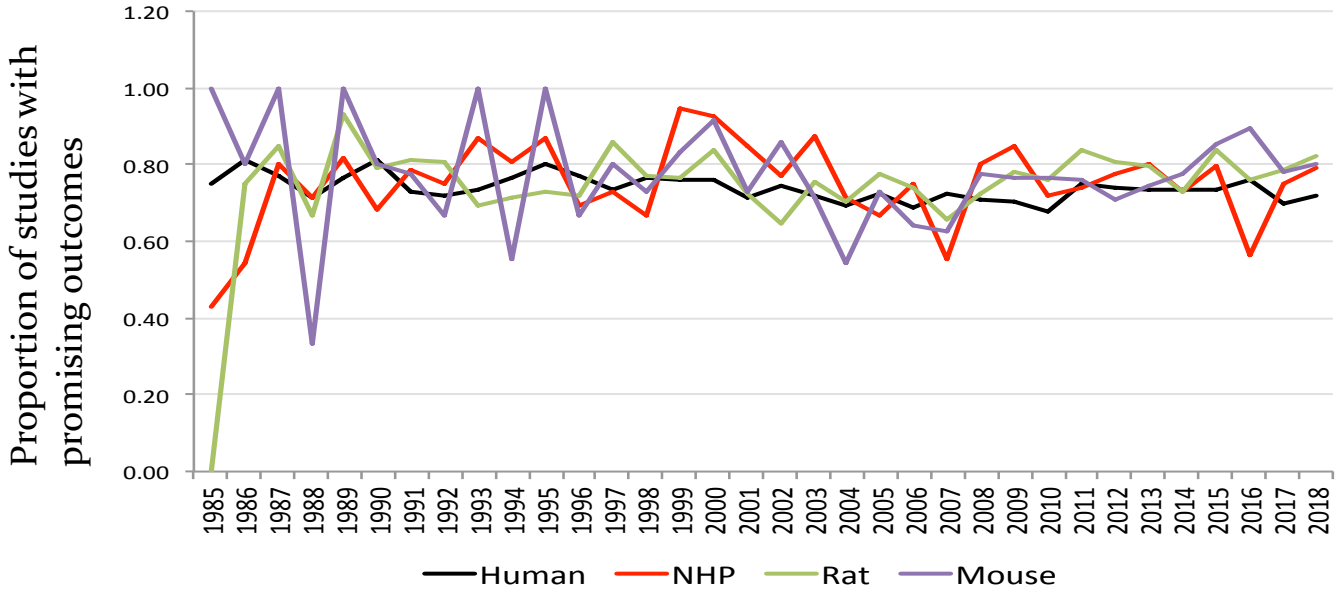
F1 score: 0.75-0.95

Validated on 14,481 abstract data set (2008–2017)

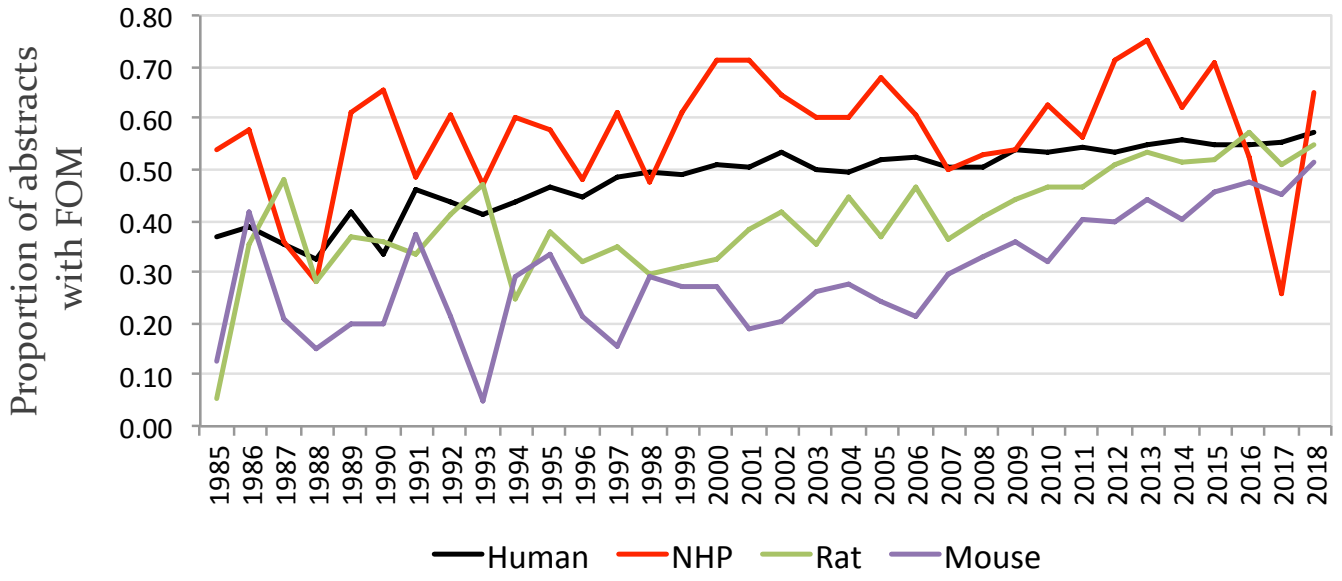
Zeiss CJ, Shin D, Vander Wyk B, Beck AP, Zatz N, Sneiderman CA, **Kilicoglu H**. Menagerie: A text-mining tool to support animal-human translation in neurodegeneration research. PLoS One. 2019

Menagerie extended to 80,000 Parkinson's Disease abstracts (1950-2019)

49 species
40 models
5000 interventions
>200 functional outcomes
>4000 genes



72-78% of studies report a promising outcome regardless of species



Functional Outcome Measures reported most often in NHPs and humans …but improving in rodents

DATASET_Interventions_Species

INTERACTIVE DATABASE: https://medicine.yale.edu/compmed/research/menagerie/

Approved for PD=1

Dopamine receptor agonists (n=414)
Levodopa/carbidopa (n=3623)
DBS (n=2798)
Selegiline (n=368)
Apomorphine (n=390)
Tolcapone (n=90)
Entacapone (n=165)
Safinamide (n=31)
Ropinirole (n=143)
Pramipexole (n=247)
Rotigotine (n=147)
Rasagiline (n=161)
Adenosine A2A receptor antagonists (n=202)
Istradefylline (n=43)
Amantadine (n=183)
Pergolide (n=140)
Pimavanserin (n=41)

Mean 29.3 years (10-35)

Not Approved for PD=0

CoQ (n=67)
Neurturin (n=42)
GDNF (n=444)
Transplantation (1985-2019, n=1100)
Neurotrophic factors (1985-2019, n= 358)
Gene therapy (1985-2019, n= 239)
NSAIDs (n= 39)
Glitazones (n=31)
Calcium channel blockers (n=33)
Antioxidant (n=262)
Glutathione (n=255)
Resveratrol (n=54)
Curcumin (n=90)
Transcranial electromagnetic stimulation (n=211
Glucocerebrosidase activators (n=22)
Nanotechnology (n= 139)
Iron chelation (n=20)
Statins (n=83)
Antihyperglycemics (n= 47)
Melatonin (n= 124)
Flavonoids (n=121)
Phenolic acids and catechols (n=92)
Estrogen (n=108)
Small interfering RNA (n=62)
Immunotherapy (n=94)
Quinones (n=50)
Heat shock response (n=29)

Mean 20.8 years (8-35)

Can we identify patterns of animal use that associate with FDA approval?

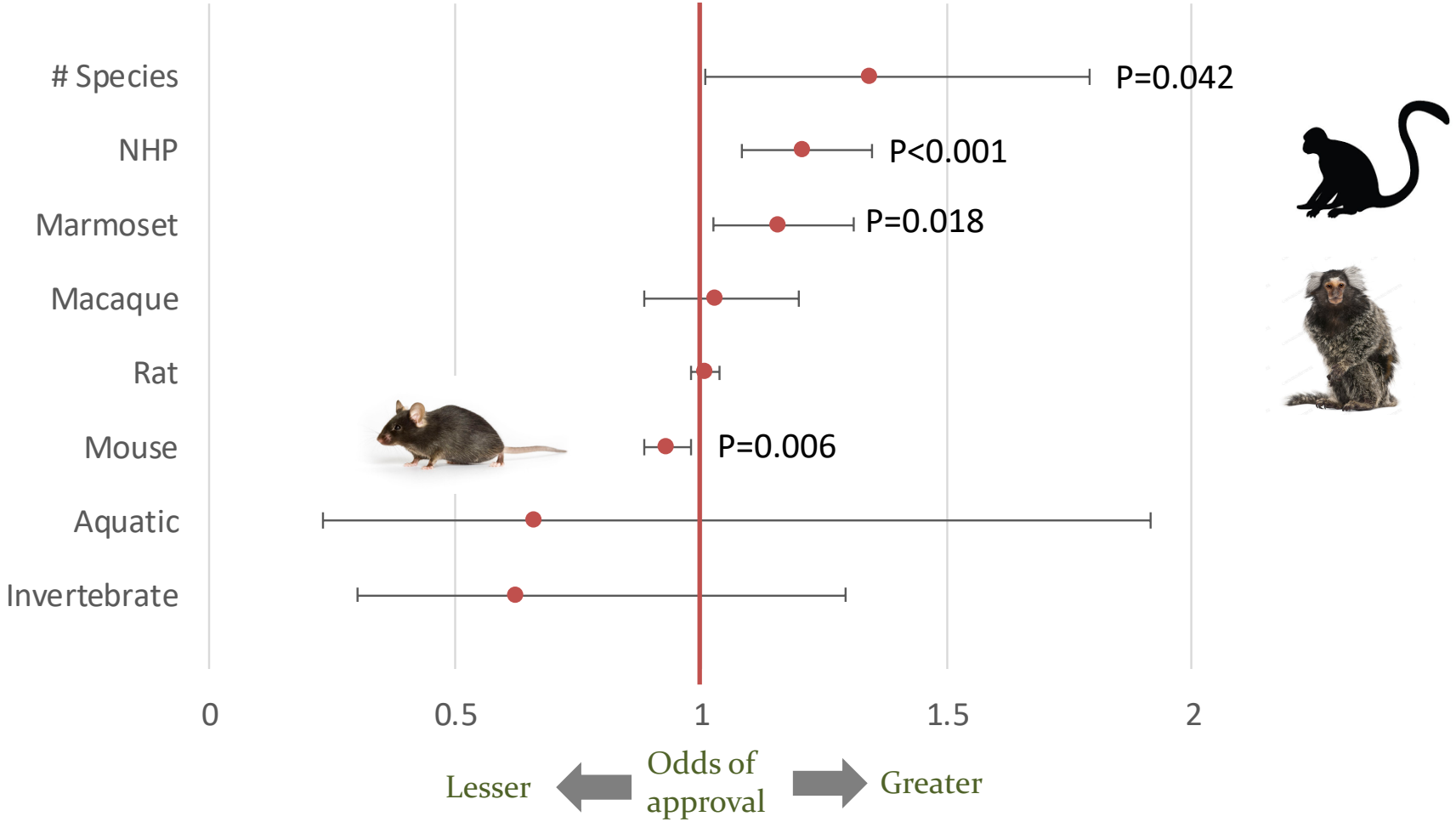
⬇

Logistic Regression Model

Bounds of Dataset_1985-2019
13, 402 abstracts
30% of DATABASE_Interventions

In those studies reporting Functional Outcome Measures AND Promising outcomes

An intervention is more likely to hold approved status if greater proportions of certain species are used

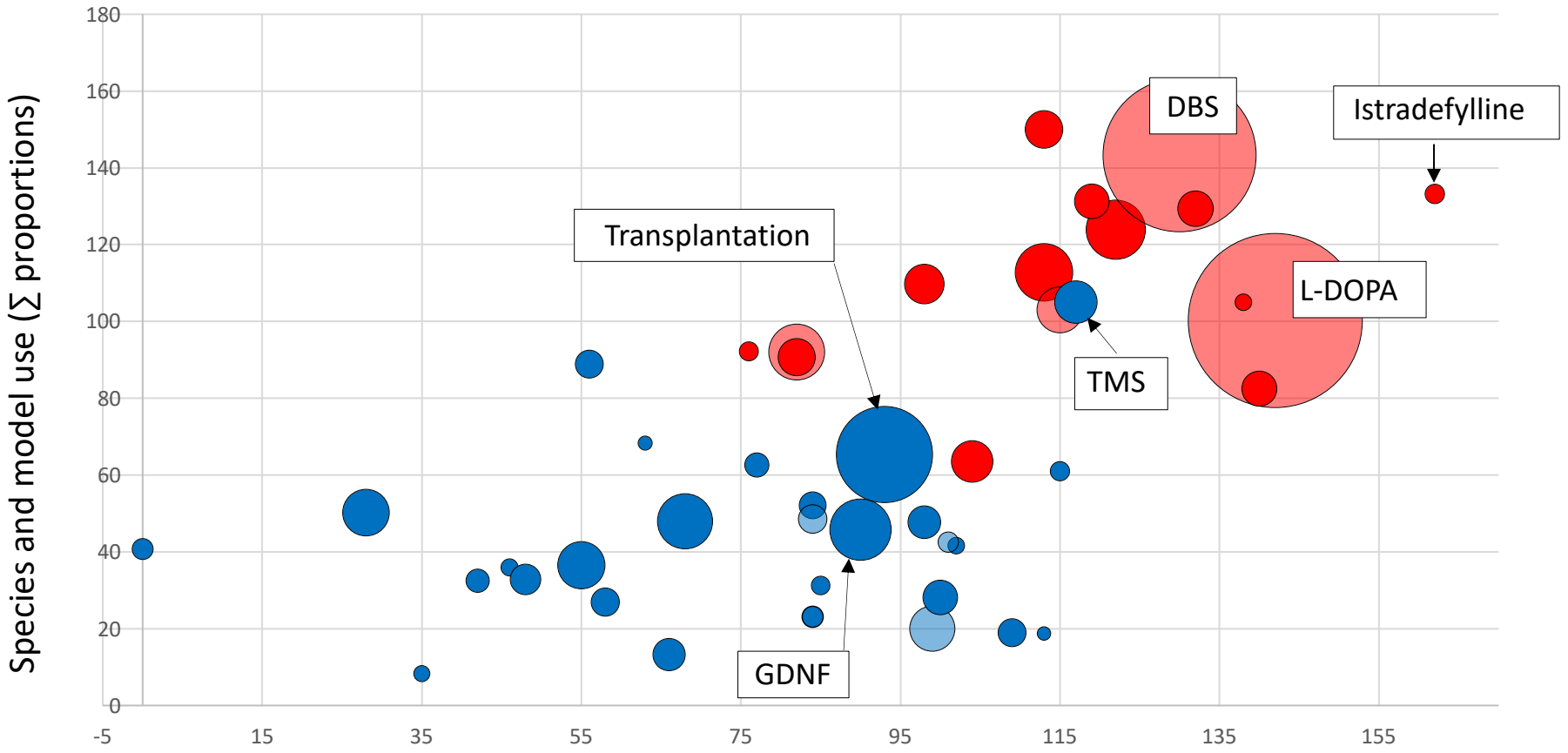


Bivariate analysis, Logistic regression, DATASET_Interventions_Species_FOM_Promising, Forest plot

Data Visualization

Species/model use vs functional outcome measure use

(Subset of variables with Odds ratio >1, p<0.05 on bivariate analysis)



Interventions: Approved (n=17); Not approved (n=27)

13, 402 of **80, 023** publications (1950-2019)

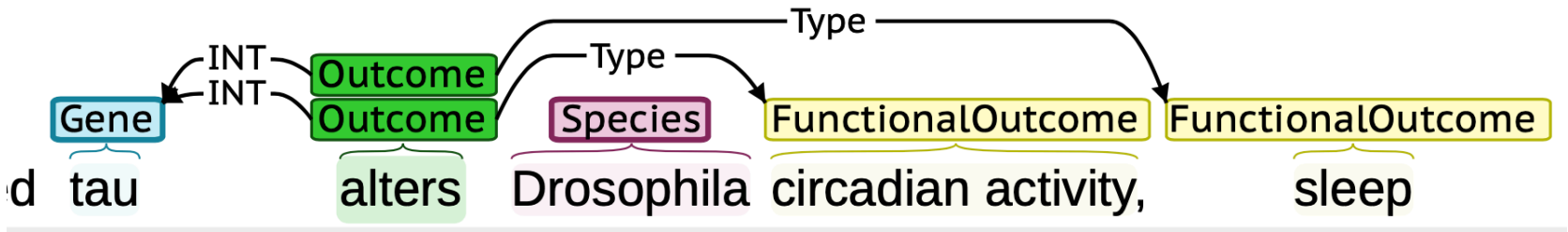
Extending Menagerie to Alzheimer's Disease: Menagerie 2.0

- Addition of new modules
  - Imaging
  - Fluid biomarkers
  - Neuropathology
  - Immunologic outcomes

- Addition of new relations
  - Between intervention and its effect on gene expression

- Addition of new rodent models
  - To include specific genetically altered rodent models

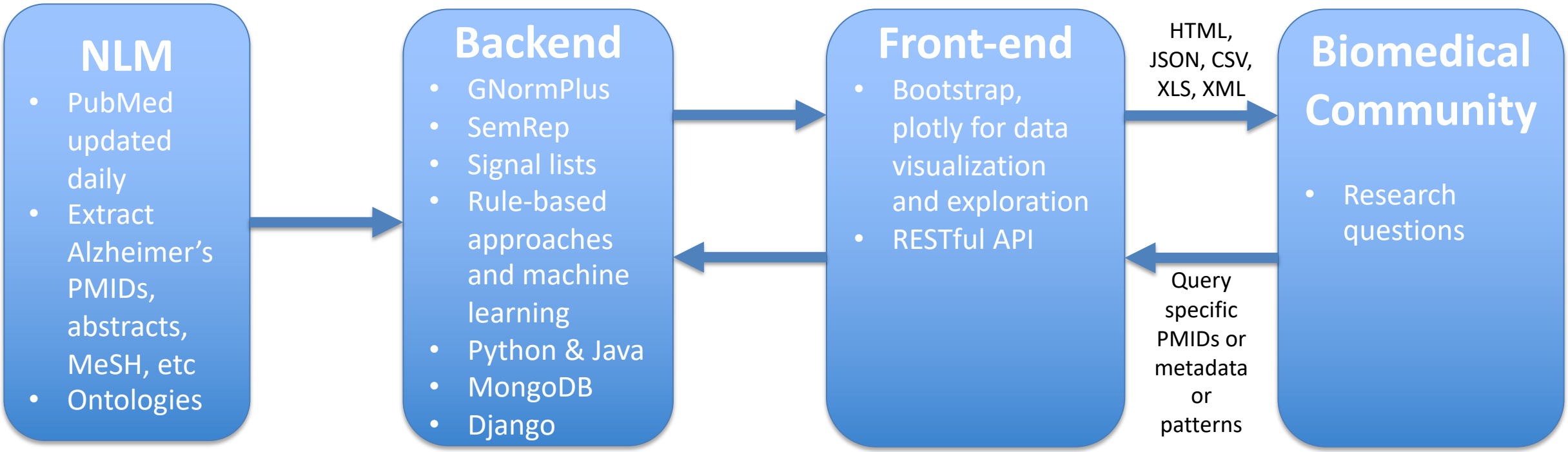
brat rapid annotation tool
v1.3 "Crunchy Frog"



Halil Kilicoglu: U. of Illinois

Menagerie 2.0

- Allow automatic updates from NLM.
- Interactive online data exploration and visualization.
- Maintain compatibility with existing Parkinson's Menagerie.
- Allow sequential capability to accommodate other neurodegenerative conditions.



Example Ontologies

MeSH

- Biomedical terms in overlapping hierarchies
- Pharmacological action and other metadata
- Humans: B01.050.150.900.649.313.988.400.112.400.400
- Often not very domain-specific (e.g. limited cell types)
- Freely downloadable

GeneOntology

- Genes
- Gene Products
- Amyloid-beta complex: GO:0106003
- Open access

National Cancer Institute thesaurus (NCIt)

- "vocabulary for clinical care, translational and basic research, and public information and administrative activities"
- Alzheimer Type II Astrocyte: C120911
- CC BY 4.0 license

UMLS

- Covers all of these and far more
- Has extensive associated infrastructure, e.g. SemRep.
- Distribution only to those with UMLS license.

- (The OBO Foundry similarly seeks to build interoperable ontologies, but it is strictly open.)

Ontologies are collections of terms, identifiers, and relationships.

e.g. in MeSH,
Alzheimer disease is a type of:

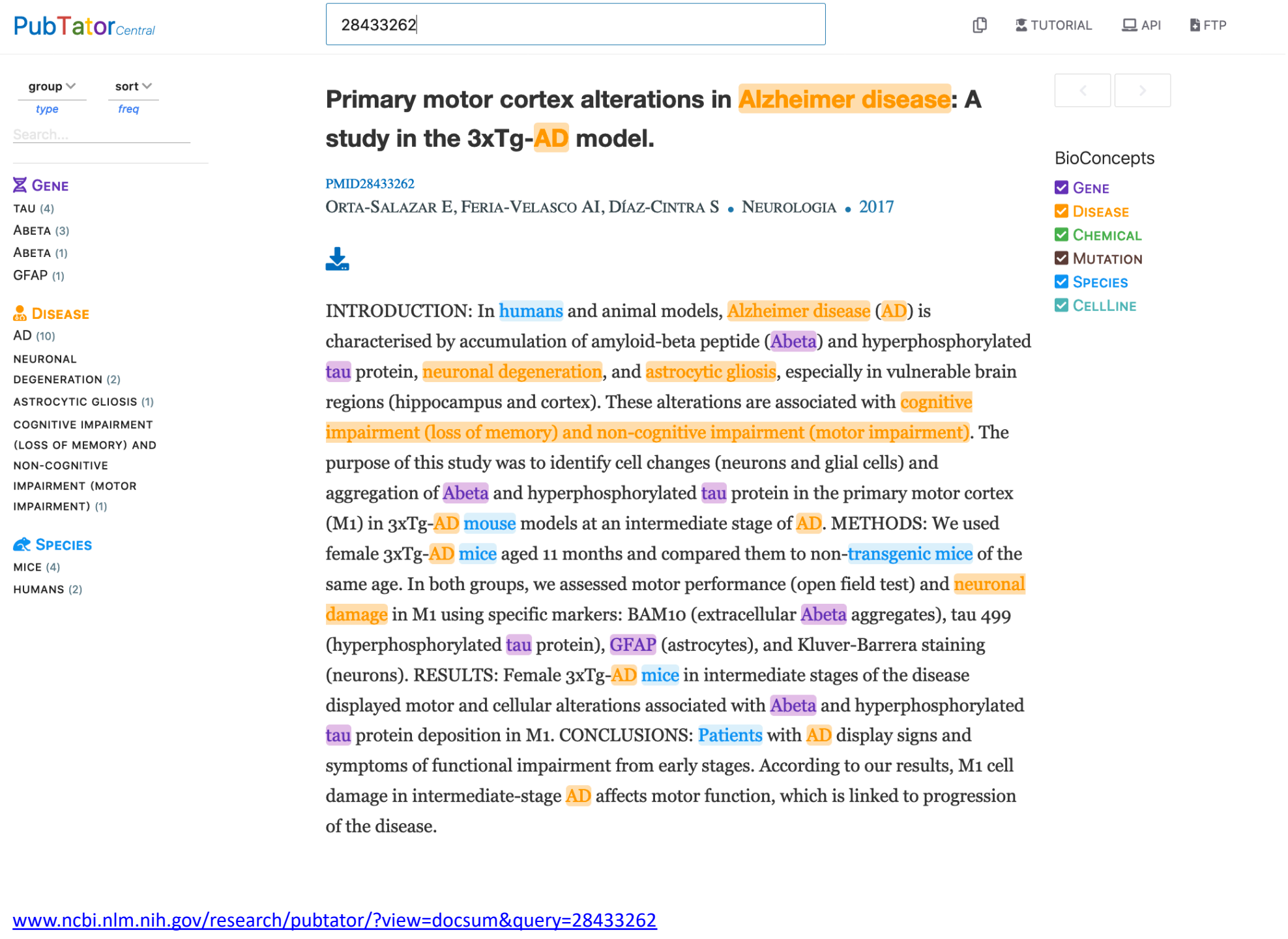
- Dementia (C10.228.140.380 and F03.615.400)
- Tauopathies (C10.574.945)

Many names:

- Alzheimer Dementia
- Alzheimer's Disease
- Alzheimer Disease
- …

Many possible qualifiers:

- Econ
- Genetics
- Immunology
- …



Context and domain specificity

- Concepts in introduction or conclusions sometimes only setting the stage (e.g. *humans*).
- Generic tools may not catch domain-specific models (e.g. *3xTg-AD mice*).

www.ncbi.nlm.nih.gov/research/pubtator/?view=docsum&query=28433262

Rules refine mentions to topics

Species and model

UMLS terms via SemRep

Pubtator/GNormPlus

Domain-specific lists

Exclusion criteria (e.g. not background)

Prefer more specific terms (e.g. mouse to rodent)

Model-species pairs based on proximity

Genes

GNormPlus, OMIM, UMLS

Prefer more specific terms

Exclusion criteria (e.g. no letters, confidence interval, ordinals, known common false-positives)

Interventions

Chemical annotation

Genes from title only

MeSH terms with pharmacologic action (PA)

UMLS with proper SemRep type info

Outcome

Tokenization, PoS tagging via Stanford CoreNLP

Bio-SCoRes for classifying and feature extraction

Two classifiers: positive and negative (allows mixed)

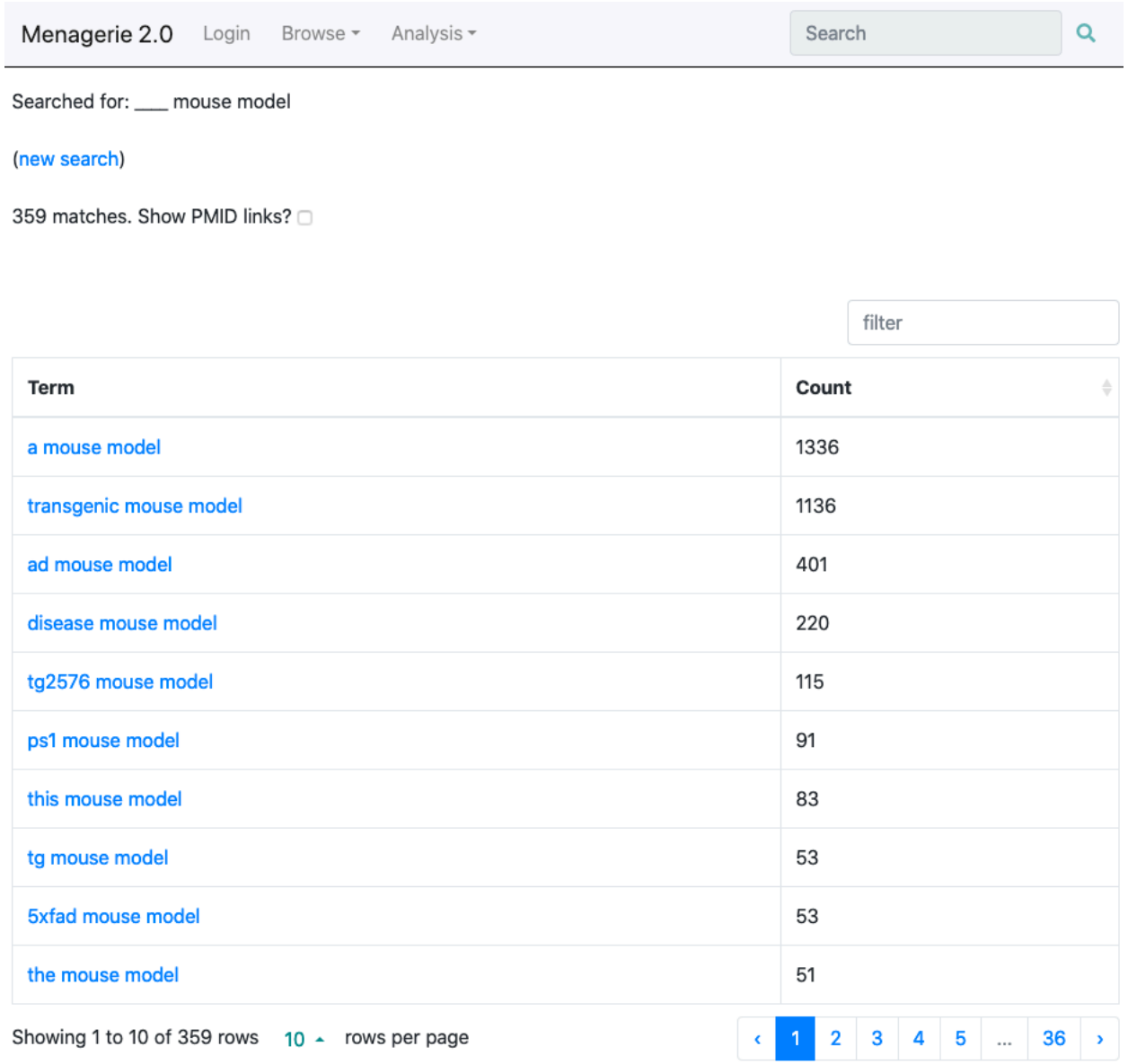
```
{
 "pmid": 29914031,
 "pmcid": null,
 "title": "Predicting … Structural Brain Biomarkers.",
 "journal": "J Alzheimers Dis. 2019",
 "pubtator": {
  "title_annotations": [{
   "id": "3",
   "infons": {"identifier": "MESH:D003072", "type": "Disease"},
   "text": "Cognitive Impairment",
   "locations": [{"offset": 72, "length": 20}]
  }, … ]},
 "abstract": "BACKGROUND: Early Alzheimer's disease (AD) …",
 "year": 2019,
 "mesh": [{"text": "Aged", "mesh_id": "D000368"}, …],
 "authors": [
  {
   "last_name": "Marizzoni",
   "first_name": "Moira",
   "initials": "M",
   "affiliation": "Laboratory of Neuroimaging and Alzheimer's…"
   }, ….],
 "doi": "10.3233/JAD-180152",
 "derived": {
  "title": {
   "animals": ["human"],
   "outcomes": []
  },…}
}
```


Data representation

- Document-oriented database (MongoDB)
  - Highly scalable.
  - Flexible schema.
  - Efficient support for multi-valued and missing fields.
- Collections
  - MeSH ontology.
  - Papers.
- Store original and derived data
  - Allows user queries for abstract patterns.
  - Easy regeneration with updated rules.
- Mappable to RDF

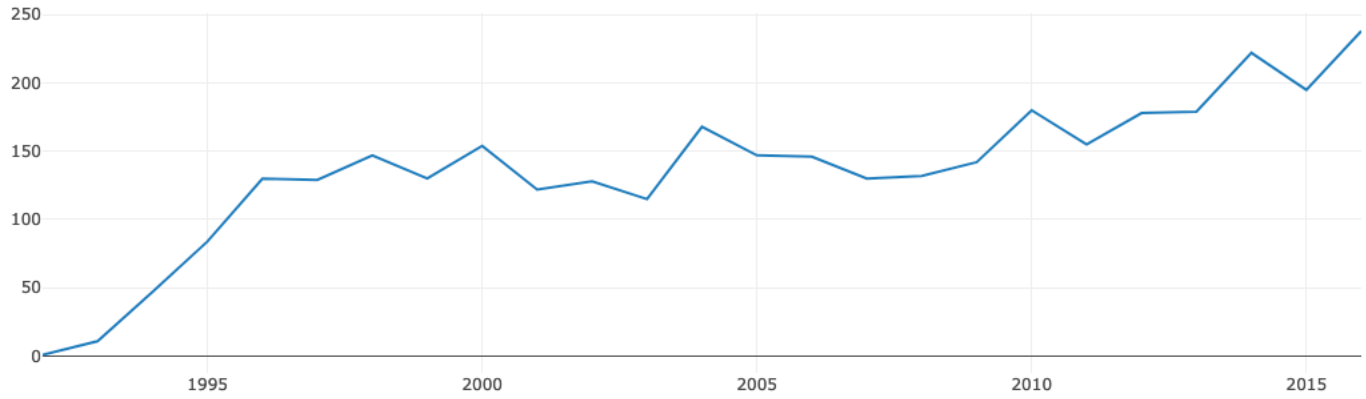
Data exploration

- Searching
  - Search for patterns
  - Search for PubMed IDs, DOIs, … (WIP)
  - Search derived metadata
- Visualization
  - Compare trends
  - Compare literature coverage (WIP)
- Cross-reference
  - E.g. what genes get studied with what model?
  - E.g. what authors look at what genes?
- Export (WIP)
  - Spreadsheet: CSV, XLS
  - JSON via API



Search results

Searched for:
Text pattern: apoe



Title	MeSH	Title Animals	Title Outcomes	Abstract Animals	Abstract Outcomes
APOE effect on Alzheimer's disease biomarkers in older adults with significant memory concern. (2015)	Adult, Aged, Aged, 80 and over, Alzheimer Disease, Amyloid beta-Peptides, Apolipoprotein E4, Biomarkers, Brain, Cognitive Dysfunction, Female, Genotype, Healthy Volunteers, Heterozygote, Humans, Magnetic Resonance Imaging, Male, Memory, Neuropsychological Tests, Positron-Emission Tomography, tau Proteins				neurodegeneration
Factors that influence the levels of cerebrospinal fluid biomarkers in memory clinic patients. (2017)	Adult, Aged, Aged, 80 and over, Alzheimer Disease, Amyloid beta-Peptides, Apolipoprotein E4, Biomarkers, Cognitive Dysfunction, Cross-Sectional Studies, Dementia, Female, Humans, Male, Middle Aged, Outpatient Clinics, Hospital, Sex	human		human	

Top MeSH terms:

- Humans (3371)
- Alzheimer Disease (3290)
- Aged (2196)
- Male (2196)
- Female (2184)
- Apolipoproteins E (2147)
- Genotype (1395)
- Apolipoprotein E4 (1324)
- Aged, 80 and over (1299)
- Middle Aged (1279)

(Total of 2806 unique MeSH terms).

Top title animals:

- human (487)
- mouse (162)
- in vitro (24)
- rat (12)
- monkey (5)
- computational model (3)
- Tg2576 mouse (3)
- APP23 mouse (2)
- drosophila (2)
- 5xFAD mouse (2)

(Total of 17 unique title animals).

Top authors:

- AD Roses (83)
- AM Saunders (73)
- H Soininen (68)
- K Blennow (64)
- L Tan (60)
- MA Pericak-Vance (57)
- DM Holtzman (57)
- J Hardy (54)
- JL Haines (51)

(Total of 12884 unique authors).

Future direction: FAIR data sharing
and the Alzheimer's data ecosystem

- We want our derived data to be FAIR:
  - **F**indable
  - **A**ccessible
  - **I**nteroperable
  - **R**eusable

- Derived data, so PURL instead of DOI.
- Mostly have ontologies from the beginning.
- RESTful API with JSON return is straightforward with our architecture.
- UMLS licensing issues.
- AlzPed interoperability.

- FAIR is about access via scripts not by humans.
- Need to understand our audience.
  - SPARQL (via Virtuoso) and RDF

```
<?xml version="1.0" ?>
…
<?properties open_access?>
…
<body id="emmm201910606-body-0001">
<sec id="emmm201910606-sec-0001">
<title>Introduction</title>
…
<fig fig-type="Figure" xml:lang="en" id="emmm201910606-fig-0005"
orientation="portrait" position="float">
<label>Figure 5</label>
<caption>
<title>The 11 novel prioritized AD risk genes are within a
lysosomal gene networks or involving innate immunity</title>
<p>Genes are depicted in their known cellular components, with
previously described established GWAS (Efthymiou &amp;amp; Goate,
<xref rid="emmm201910606-bib-0010" ref-type="ref">2017</xref>;
Verheijen &amp;amp; Sleegers, <xref rid="emmm201910606-bib-0062"
ref-type="ref">2018</xref>) in black and the prioritized novel
genes (this paper) in red. See discussion section for further
functional annotation and references.</p>
</caption>
<graphic id="nlm-graphic-15" xlink:href="EMMM-12-e10606-g009"/>
</fig>
…
```

Future direction: mining full-text

Two general challenges with mining full-text:

- Licensing issues
- Scope – e.g. background vs subject

Most PubMed Central articles may be read freely but do not have a permissive license.

However, for Alzheimer's about 40% have a CC-BY license.

These are available in a structured format (left) where introduction, figure captions, methods, etc can readily be separated.

Sierksma, A., Lu, A., Mancuso, R., Fattorelli, N., Thrupp, N., Salta, E., ... & Fiers, M. (2020). Novel Alzheimer risk genes determine the microglia response to amyloid-β but not to TAU pathology. *EMBO Molecular Medicine*, 12(3), e10606.
XML from: eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=pmc&retmode=xml&id=7059012

Thank you

- Ongoing annotation
  - Sami Elzrasky
  - Rory Edwards
- Prototype interfaces
  - Cameron Conte
  - Evan Cudone
- Yale ADRC
- NIH P30 AG066508