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

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

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

Translatability vs Reproducibility

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

Menagerie


F1 score: 0.75-0.95

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

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

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

DATASET_Interventions_Species

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

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)
Prampipexole (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)
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
Use of functional outcome measures (∑ proportions)

Species and model use (∑ proportions)

(interventional 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

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.

**NLM**
- PubMed updated daily
- Extract Alzheimer’s PMIDs, abstracts, MeSH, etc
- Ontologies

**Backend**
- GNormPlus
- SemRep
- Signal lists
- Rule-based approaches and machine learning
- Python & Java
- MongoDB
- Django

**Front-end**
- Bootstrap, plotly for data visualization and exploration
- RESTful API

**Biomedical Community**
- Research questions

**Query specific PMIDs or metadata or patterns**

**HTML, JSON, CSV, XLS, XML**
# Example Ontologies

<table>
<thead>
<tr>
<th>MeSH</th>
<th>GeneOntology</th>
<th>National Cancer Institute thesaurus (NCIt)</th>
<th>UMLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Biomedical terms in overlapping hierarchies</td>
<td>• Genes</td>
<td>• “vocabulary for clinical care, translational and basic research, and public information and administrative activities”</td>
<td>• Covers all of these and far more</td>
</tr>
<tr>
<td>• Pharmacological action and other metadata</td>
<td>• Gene Products</td>
<td>• Alzheimer Type II Astrocyte: C120911</td>
<td>• Has extensive associated infrastructure, e.g. SemRep.</td>
</tr>
<tr>
<td>• Humans: B01.050.150.900.649.313.988.400.112.400.400</td>
<td>• Amyloid-beta complex: GO:0106003</td>
<td></td>
<td>• Distribution only to those with UMLS license.</td>
</tr>
<tr>
<td>• Often not very domain-specific (e.g. limited cell types)</td>
<td>• Open access</td>
<td></td>
<td>(The OBO Foundry similarly seeks to build interoperable ontologies, but it is strictly open.)</td>
</tr>
<tr>
<td>• Freely downloadable</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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
• ...
Primary motor cortex alterations in Alzheimer disease: A study in the 3xTg-AD model.

INTRODUCTION: In humans and animal models, Alzheimer disease (AD) is characterised by accumulation of amyloid-beta peptide (Abeta) and hyperphosphorylated tau protein, neuronal degeneration, and astrocytic gliosis, especially in vulnerable brain regions (hippocampus and cortex). These alterations are associated with cognitive impairment (loss of memory) and non-cognitive impairment (motor impairment). The purpose of this study was to identify cell changes (neurons and glial cells) and aggregation of Abeta and hyperphosphorylated tau protein in the primary motor cortex (M1) in 3xTg-AD mouse models at an intermediate stage of AD. METHODS: We used female 3xTg-AD mice aged 11 months and compared them to non-transgenic mice of the same age. In both groups, we assessed motor performance (open field test) and neuronal damage in M1 using specific markers: BAM10 (extracellular Abeta aggregates), tau 499 (hyperphosphorylated tau protein), GFAP (astrocytes), and Kluver-Barrera staining (neurons). RESULTS: Female 3xTg-AD mice in intermediate stages of the disease displayed motor and cellular alterations associated with Abeta and hyperphosphorylated tau protein deposition in M1. CONCLUSIONS: Patients with AD display signs and symptoms of functional impairment from early stages. According to our results, M1 cell damage in intermediate-stage AD affects motor function, which is linked to progression of the disease.
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)
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
Future direction: FAIR data sharing
and the Alzheimer’s data ecosystem

- We want our derived data to be FAIR:
  - Findable
  - Accessible
  - Interoperable
  - Reusable

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

- 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.
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.
Thank you

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