Knowledge-aware automated machine learning

with application to Alzheimer's disease risk prediction

Jason H. Moore, PhD, FACMI, FIAHSI, FASA
Chair, Department of Computational Biomedicine
Director, Center for Artificial Intelligence Research and Education
Cedars-Sinai Medical Center

Adjunct Professor of Biostatistics
Affiliate Professor of Bioengineering
UCLA

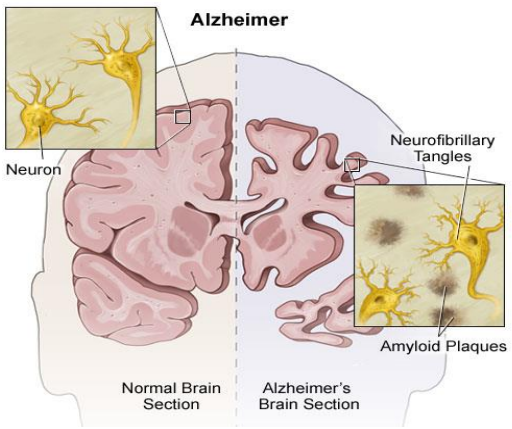
jason.moore@csmc.edu

Cedars Sinai

cedars-sinai.org

1

Alzheimer's Disease

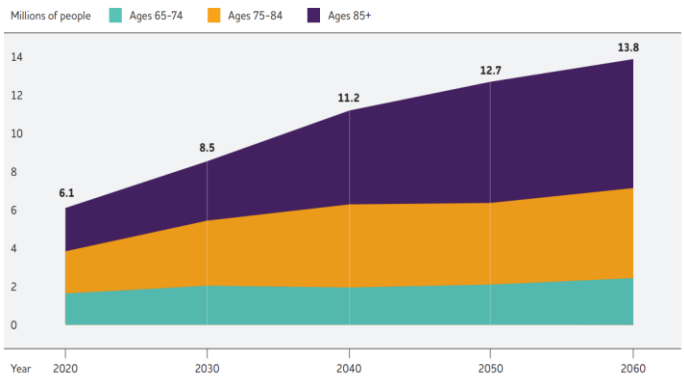


- Impaired thinking, and behavior
- Confusion
- Restlessness
- Personality and behavior changes
- Impaired judgment
- Impaired communication
- Inability to follow directions
- Language deterioration

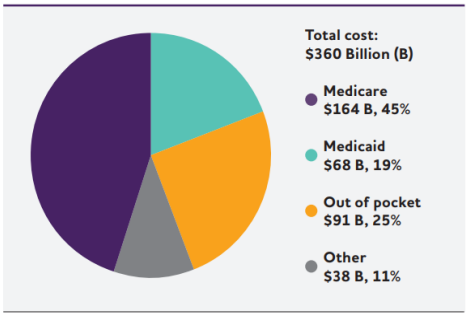
2

Alzheimer's Disease Cases are Increasing

Projected Number of People Age 65 and Older (Total and by Age) in the U.S. Population with Alzheimer's Dementia, 2020 to 2060



Distribution of Aggregate Costs of Care by Payment Source for Americans Age 65 and Older with Alzheimer's or Other Dementias, 2024*



Cedars Sinai

https://www.alz.org/media/Documents/alzheimers-facts-and-figures.pdf 3

3

NIH Funding for Alzheimer's Disease Research

AIM ALZHEIMER'S IMPACT MOVEMENT ALZHEIMER'S ASSOCIATION

Alzheimer's and Related Dementia Research Funding at the NIH

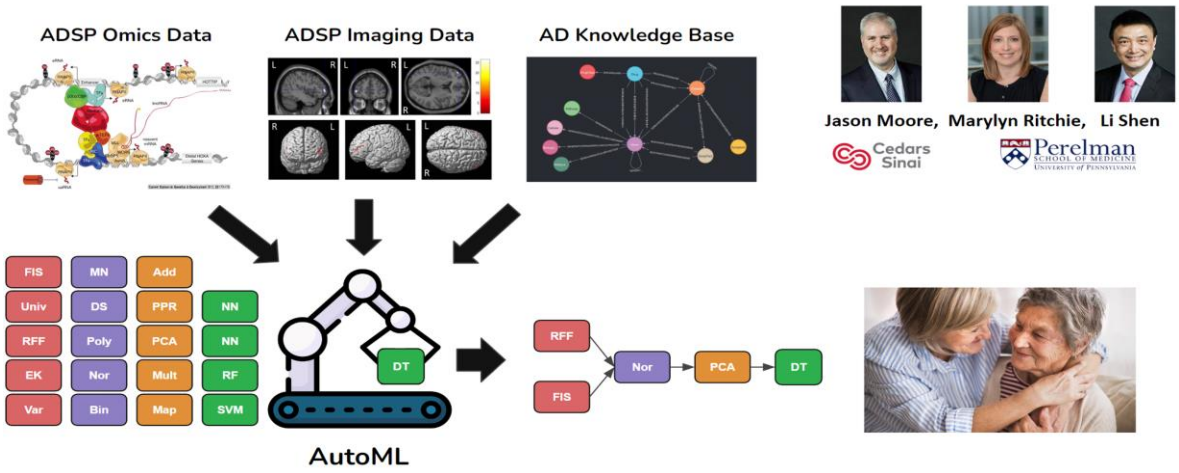
Source: NIH [https://report.nih.gov/funding/categorical-spending] with the 2024 figure reflecting NIH's estimated 2023 funding plus the additional increase of $100M appropriated through the FY2024 funding bill.

Cedars Sinai

https://alzimpact.org/research 4

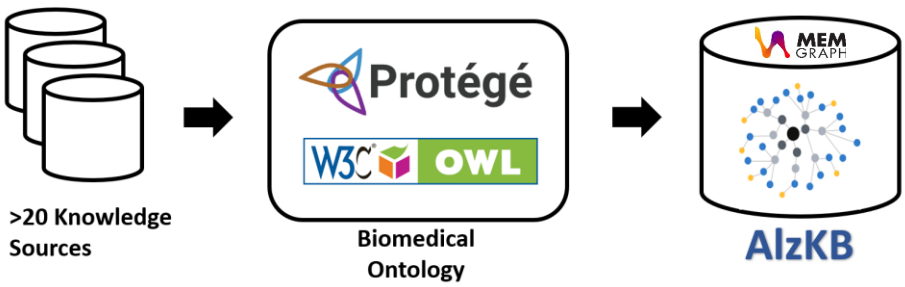
4

Risk Prediction and Drug Discovery for Alzheimer's Disease



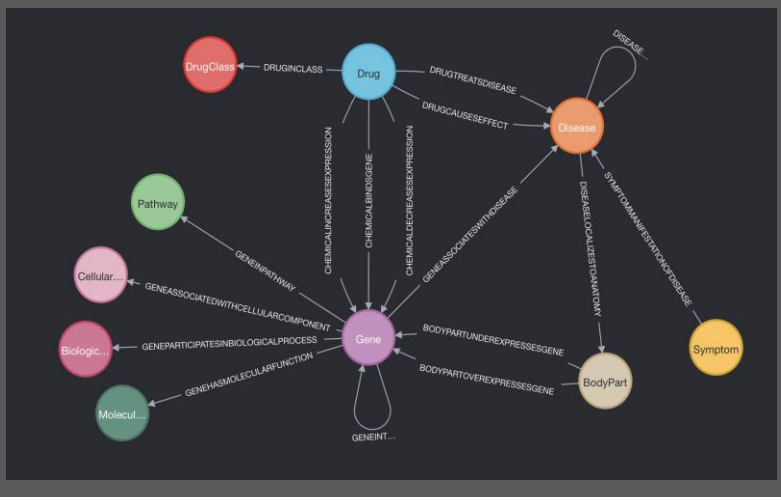
5

The Alzheimer's Disease Knowledgebase (AlzKB.ai)

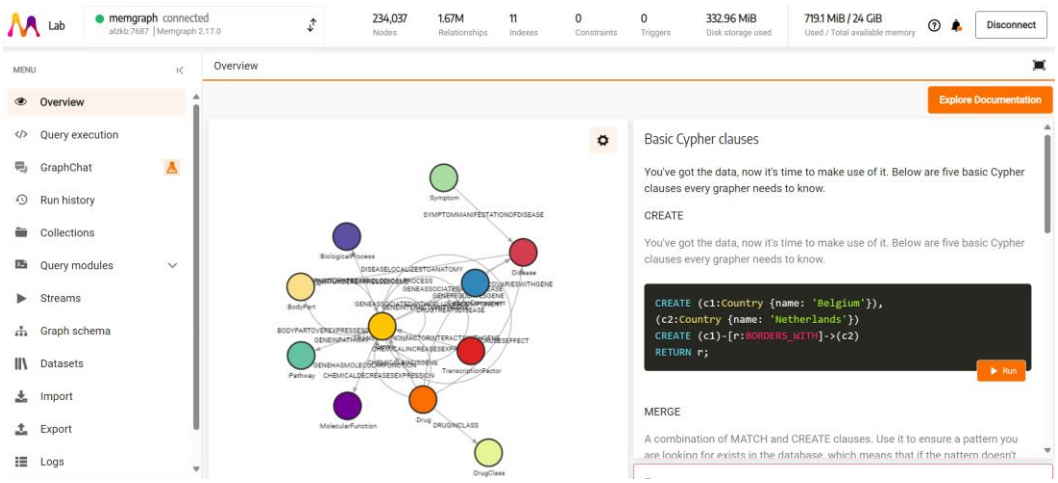


6

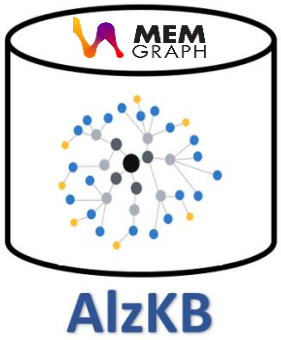
The Alzheimer's Disease Knowledgebase (AlzKB.ai)



AlzKB.AI in Memgraph



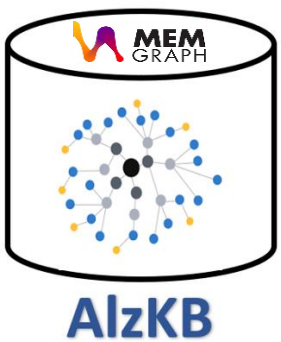
Querying AlzKB using the Cypher language




```
MATCH p=shortestPath(
  (a:Item {itemId:'Q332'})-[*]->(b:Item {itemId:'Q35120'})
)
WHERE NONE(x IN RELATIONSHIPS(p)
  WHERE (x.propId <> 'P279') AND
    (x.propId <> 'P31') AND
    (x.propId <> 'P361')
)
RETURN p
```

```
MATCH (a:Item), (b:Item)
WHERE a.itemId IN ['Q2', 'Q405', 'Q525']
  AND b.itemId IN ['Q2', 'Q405', 'Q525']
WITH a, b
OPTIONAL MATCH (a)-[rel]-(b)
RETURN a, b, collect(rel)
```


Can we use an LLM to query AlzKB?



OpenAI Chat GPT-4

No! Accuracy = 0.5 to 0.55

KRAGEN

Bioinformatics, 2024, **40(6)**, btae353
https://doi.org/10.1093/bioinformatics/btae353
Advance Access Publication Date: 3 June 2024
Applications Note

Bioinformatics OXFORD

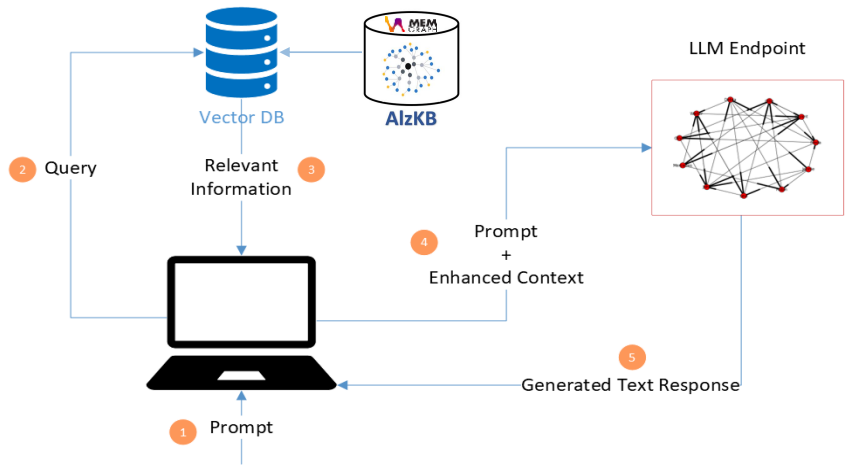
Databases and ontologies

KRAGEN: a knowledge graph-enhanced RAG framework for biomedical problem solving using large language models

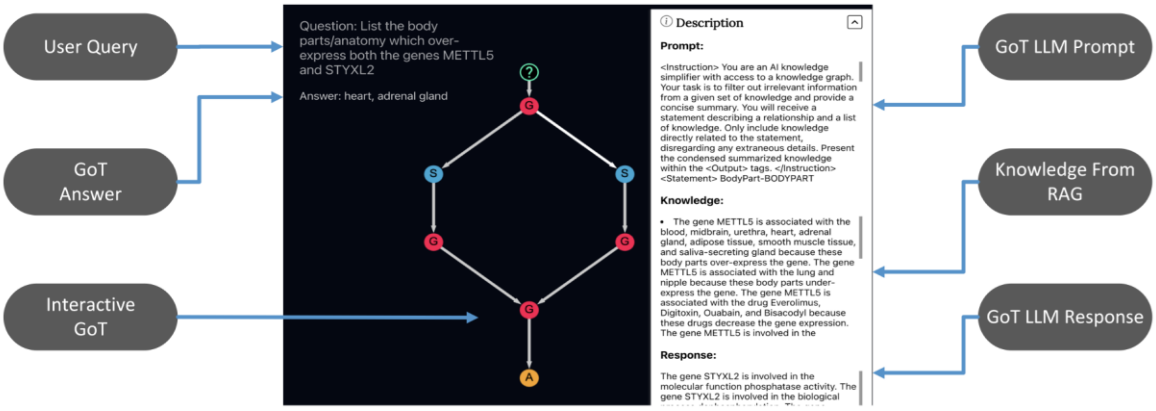
Nicholas Matsumoto[1], Jay Moran[1], Hyunjun Choi[1], Miguel E. Hernandez[1], Mythreye Venkatesan[1], Paul Wang[1], Jason H. Moore [1,*]

[1]Department of Computational Biomedicine, Center for Artificial Intelligence Research and Education, Cedars Sinai Medical Center, West Hollywood, CA 90069, United States

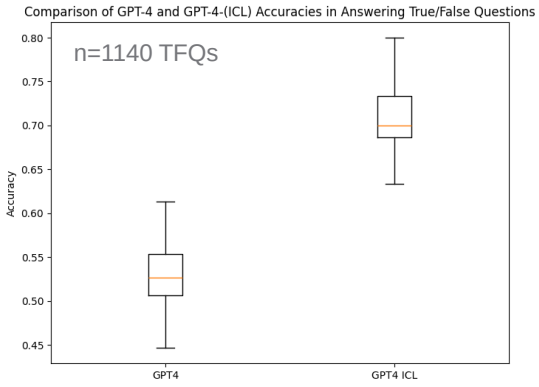
Querying AlzKB using LLMs with In Context Learning (ICL)

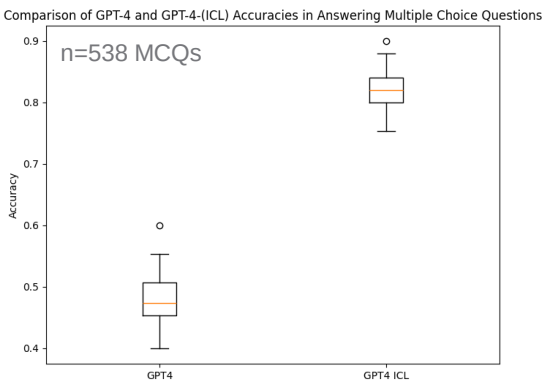


Deconstructing queries with Graph of Thoughts (GoT)

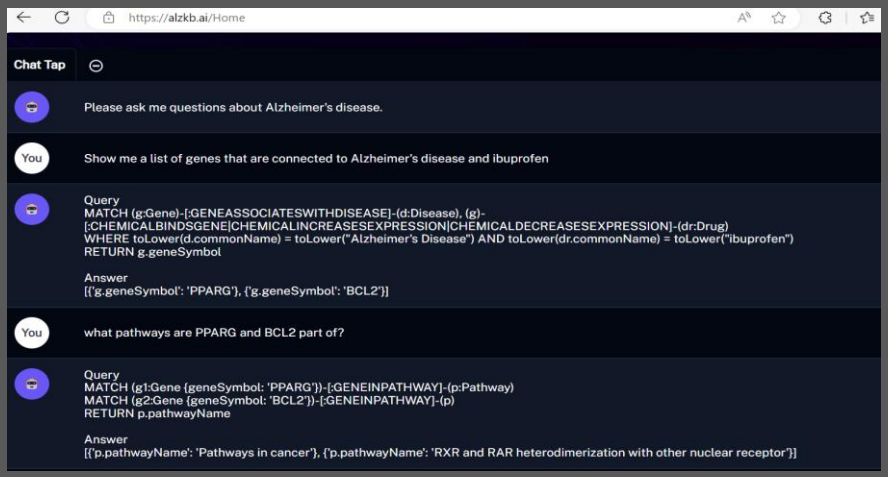


Querying AlzKB using LLMs with ICL, RAG, and GoT

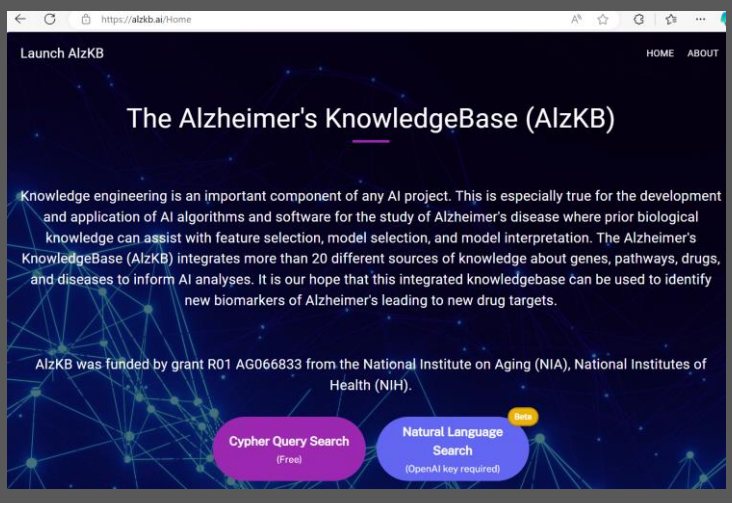




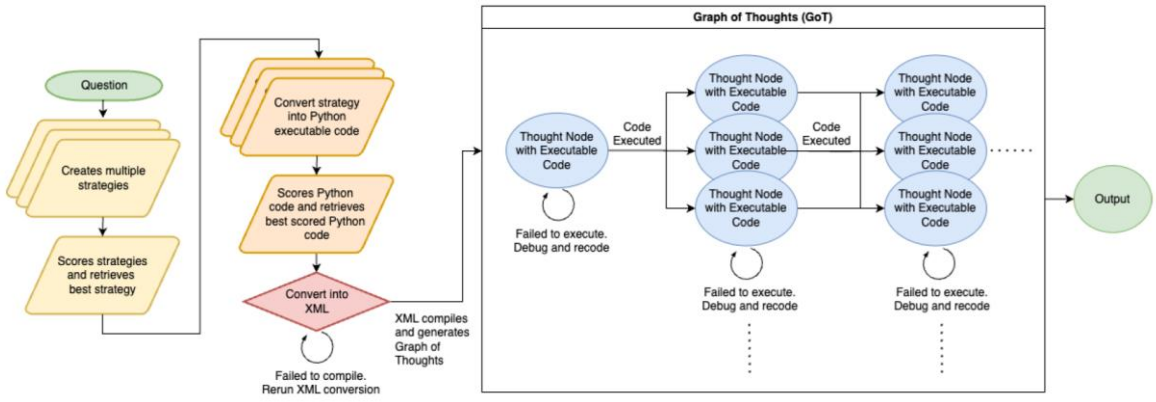
Querying AlzKB using KRAGEN



Querying AlzKB using KRAGEN



ESCARGOT: Dynamic Graph of Thoughts



ESCARGOT (in revision)

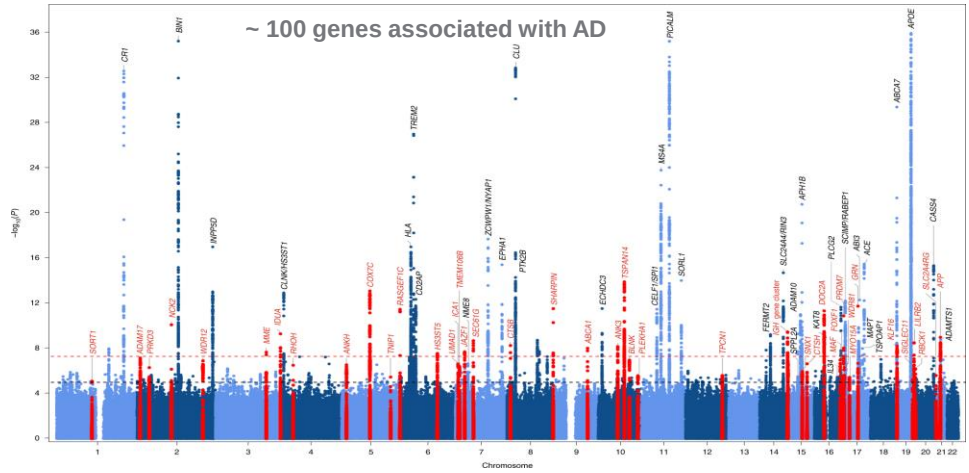
ESCARGOT: An AI Agent Leveraging Large Language Models, Dynamic Graph of Thoughts, and Biomedical Knowledge Graphs for Enhanced Reasoning

Nicholas Matsumoto,[1] Hyunjun Choi,[1] Jay Moran,[1] Miguel E. Hernandez,[1] Mythreye Venkatesan,[1] Xi Li,[1] Jui-Hsuan Chang,[1] Paul Wang[1] and Jason H. Moore[1,*]

[1]Department of Computational Biomedicine, Center for Artificial Intelligence Research and Education, Cedars Sinai Medical Center, 700 N. San Vicente Blvd., Pacific Design Center, Suite G-541H, West Hollywood, 90069, CA, USA

Dataset	GPT-4o Mini	RAG	ESCARGOT
Openended 1-hop (508 questions)	4.2%	55.5%	**88.4%**
Openended 2-hop (450 questions)	4.9%	23.1%	**85.8%**
True/False 1-hop (560 questions)	60.5%	85.2%	**90.9%**
True/False 2-hop (540 questions)	59.4%	**75.6%**	74.1%
Multiple Choice 1-hop (498 questions)	58.2%	88.8%	**93.4%**
Multiple Choice 2-hop (419 questions)	61.3%	86.4%	**89.5%**

Genetic Analysis of Alzheimer's Disease



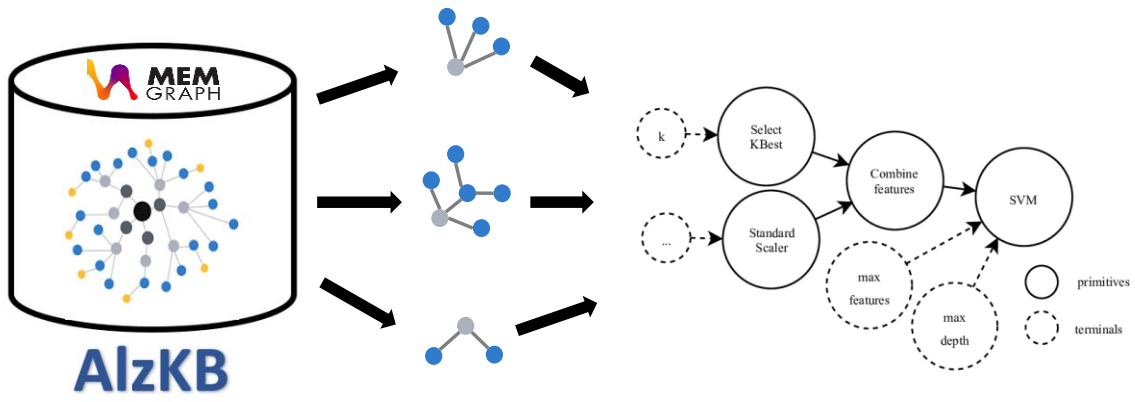
Cedars Sinai

Bellenguez et al. *Nature Genetics* (2024) https://www.nature.com/articles/s41588-022-01024-z 19

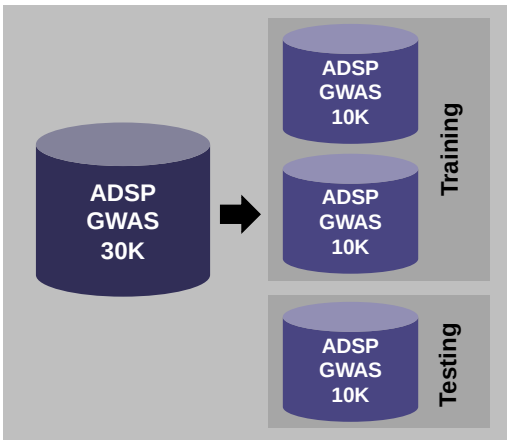
19

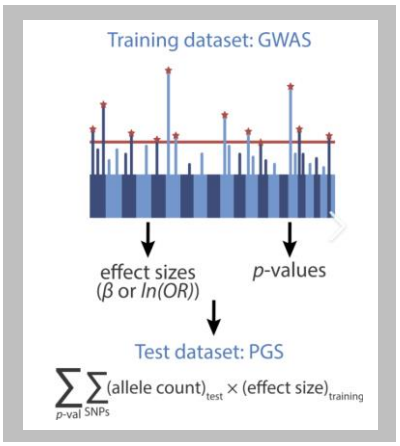
20

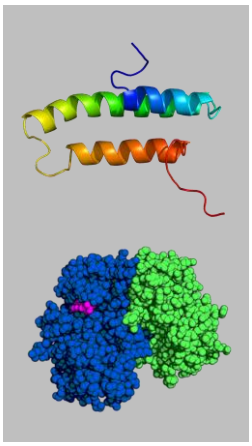
Automated Feature Selection and Machine Learning



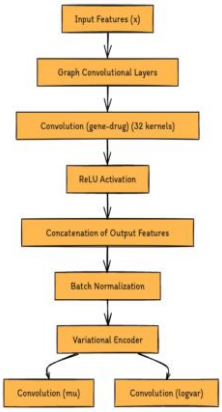
Data Preparation (n=30k)

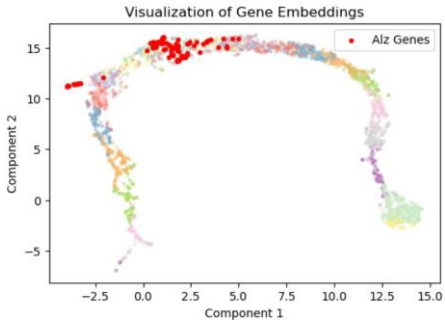


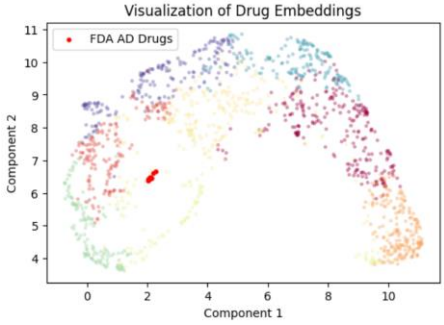




Defining Distance from Alzheimer's Using a Variational Graph Auto Encoder

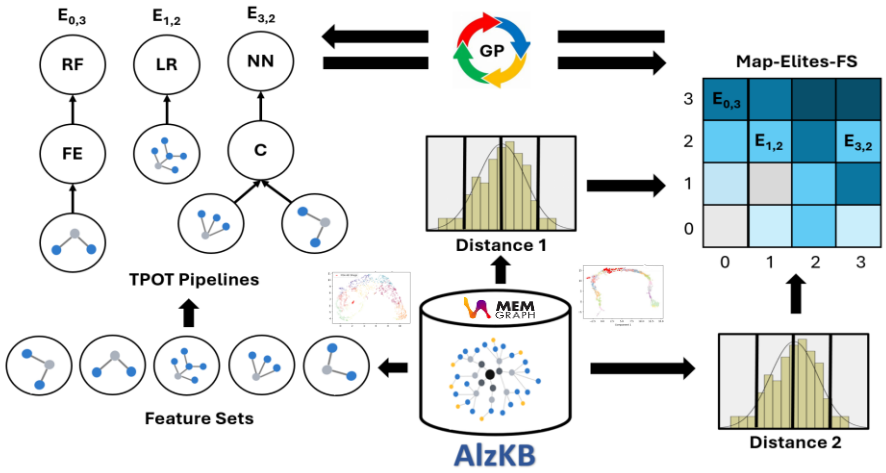






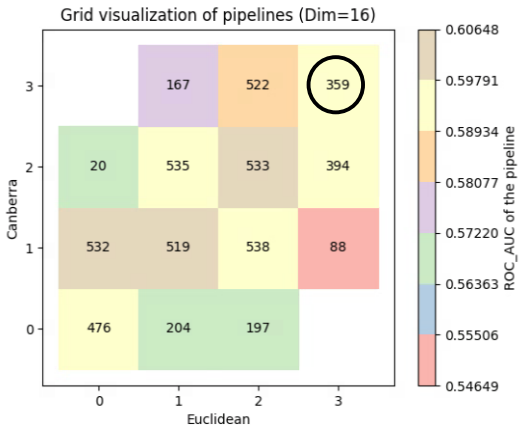
Cedars Sinai

Knowledge-Guided Automated Machine Learning Using Map-Elites



Cedars Sinai

Knowledge-Guided Automated Machine Learning Using Map-Elites



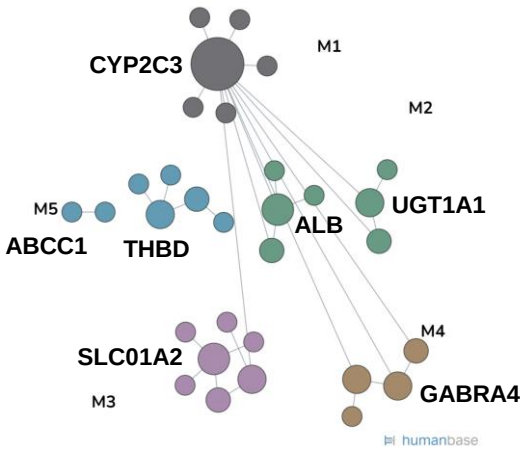
Feature set 359

GABRB1 GABRR1 GABRA5 GABRE GABRR3 GABRA3 GABRQ GABRG2 GABRB3 GABRA1 GABRD GABRG3 GABRA2 GABRP GABRR2 CYP2C8 GABRG1 CYP2B6 GABRA4 CYP2C19 GABRB2 TSPO CYP2C9 GABRA6 CYP3A4 UGT1A3 ALB ABCC4 THBD ABCC1 SLCO2B1 UGT2B7 PTGS1 CYP2C9 CFTR CXCL8 PPARG SLC22A11 ABCB1 PLAT UGT1A1 SLCO1A2 SLC22A8 CYP2C19 UGT2B4 BCL2 UGT1A9 SLC22A6 PTGS2 CYP2C8

Drugs

Temazepam (a benzodiazepine)
Ibuprofen

Gene Interpretation Using HumanBase



M1: Drug Metabolism

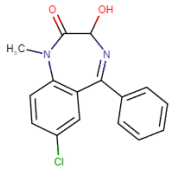
M2: Drug Metabolism

M3: Chloride Transport

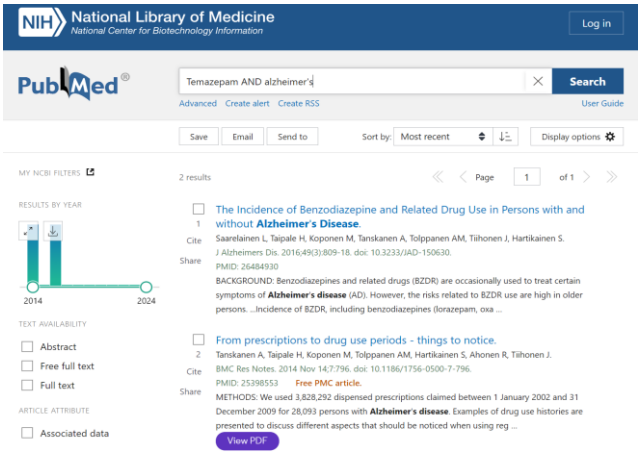
M4: Chloride Transport

M5: Receptor Metabolism

Drug Interpretation



Temazepam, a benzodiazepine, targets the GABA neurotransmitter



The Digital Data Science Assistant

Acknowledgements

- TPOT: Jun Choi, Nick Matsumoto, Jay Moran, Pedro Ribeiro, Jose Hernandez
- AlzKB: Joe Romano
- KRAGEN & ESCARGOT: Jun Choi, Nick Matsumoto, Jay Moran, Miguel Hernandez
- Map-Elites-FS: Sisi Shao
- NIA U01 AG066833, R01 LM010098, R01 LM014572

- https://github.com/epistasislab/TPOT
- https://github.com/epistasislab/KRAGEN
- https://github.com/epistasislab/ESCARGOT
- https://github.com/epistasislab/ALZKB

Questions?