

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

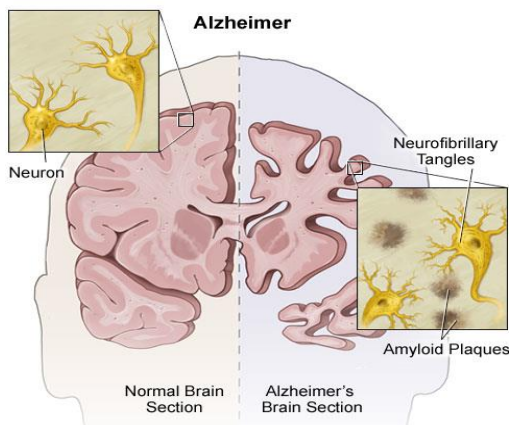
jason.moore@csmc.edu



cedars-sinai.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



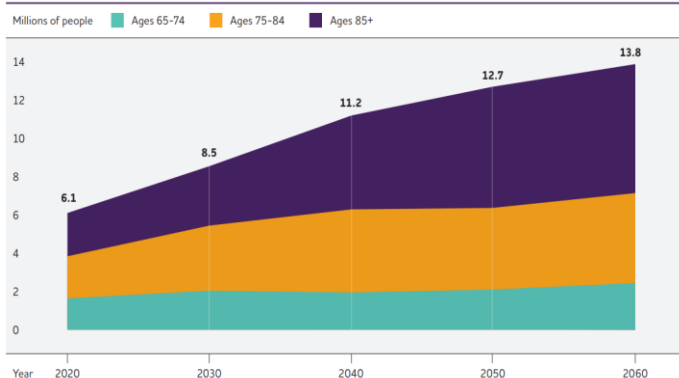
<https://stanfordhealthcare.org/medical-conditions/brain-and-nerves/alzheimers-disease.html>

2

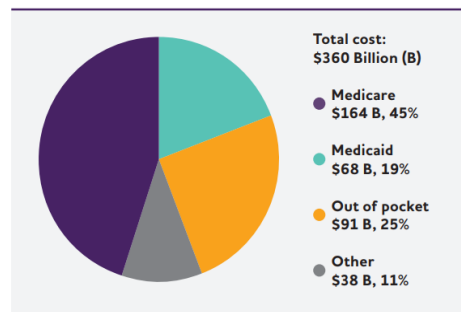
2

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*



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

3

NIH Funding for Alzheimer's Disease Research



Alzheimer's and Related Dementia Research Funding at the NIH



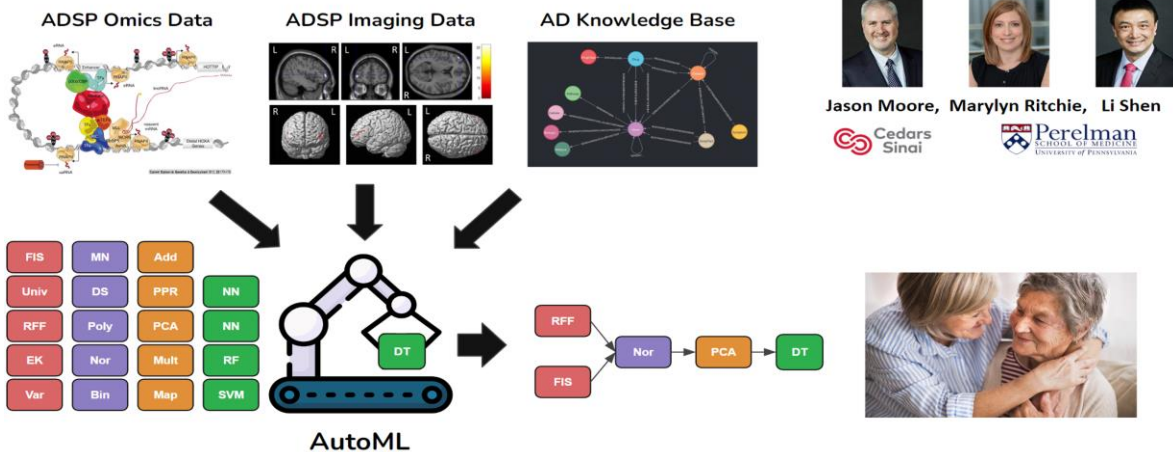
Source: NIH (https://report.nih.gov/funding/categorical_spending/) with the 2024 figure reflecting NIH's estimated 2023 funding plus the additional increase of \$100M appropriated through the FY2024 funding bill.



<https://alzimpact.org/research> 4

4

Risk Prediction and Drug Discovery for Alzheimer's Disease



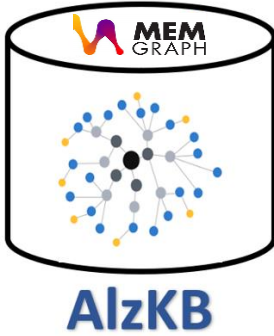
5

The Alzheimer's Disease Knowledgebase (AlzKB.ai)



6

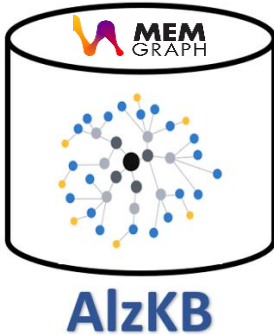
Querying AlzKB using the Cypher language



```
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 Matsumoto¹, Jay Moran¹, Hyunjun Choi¹, Miguel E. Hernandez¹, Mythreye Venkatesan¹, Paul Wang¹, Jason H. Moore^{1,*}

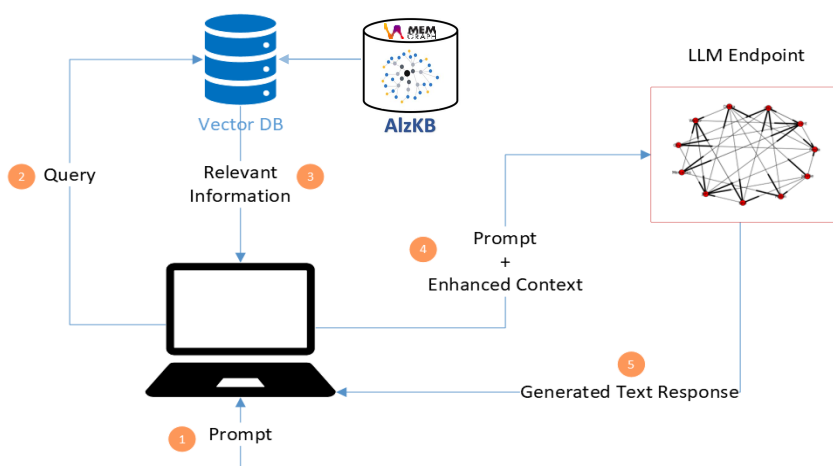
¹Department of Computational Biomedicine, Center for Artificial Intelligence Research and Education, Cedars Sinai Medical Center, West Hollywood, CA 90069, United States



Matsumoto et al., *Bioinformatics*, 2024, 40(6), btae353 11

11

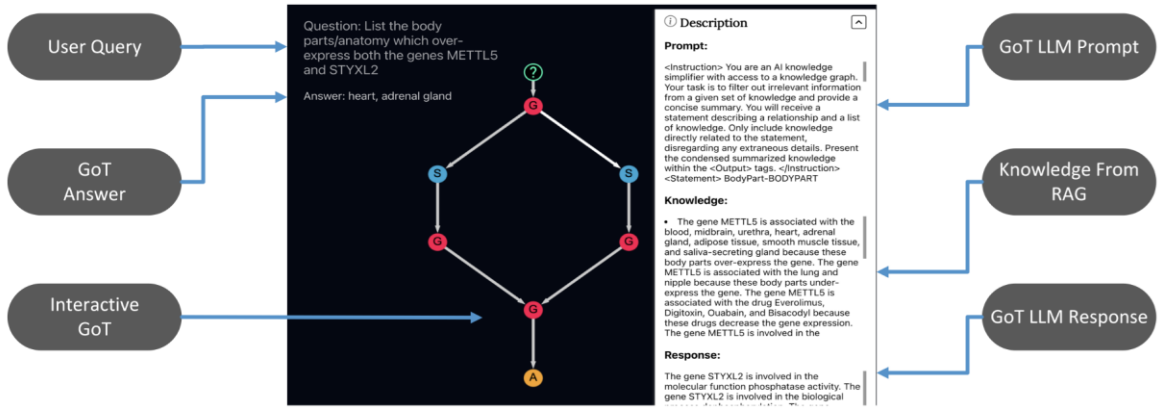
Querying AlzKB using LLMs with In Context Learning (ICL)



Matsumoto et al., *Bioinformatics*, 2024, 40(6), btae353 12

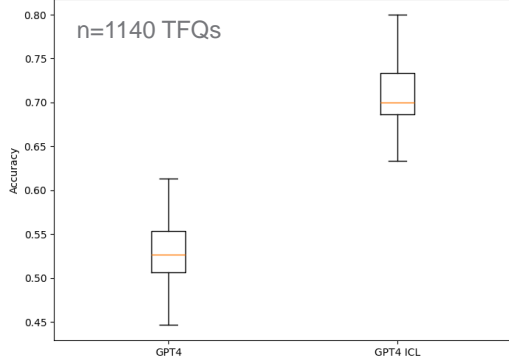
12

Deconstructing queries with Graph of Thoughts (GoT)

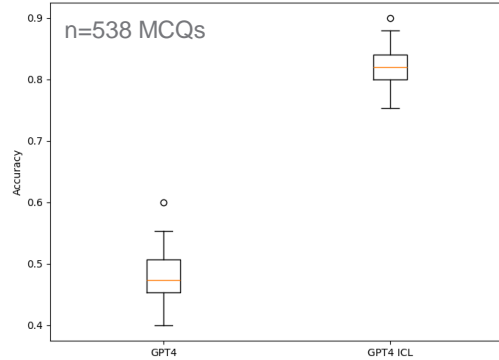


Querying AlzKB using LLMs with ICL, RAG, and GoT

Comparison of GPT-4 and GPT-4-(ICL) Accuracies in Answering True/False Questions



Comparison of GPT-4 and GPT-4-(ICL) Accuracies in Answering Multiple Choice Questions



Querying AlzKB using KRAGEN

Chat Tap

Please ask me questions about Alzheimer's disease.

You: Show me a list of genes that are connected to Alzheimer's disease and ibuprofen

Query
`MATCH (g:Gene)-[:GENEASSOCIATESWITHDISEASE]-(:d:Disease), (g)-[:CHEMICALBINDSGENE|CHEMICALINCREASESEXPRESSION|CHEMICALDECREASESEXPRESSION]-(:dr:Drug)
 WHERE toLower(d.commonName) = toLower("Alzheimer's Disease") AND toLower(dr.commonName) = toLower("ibuprofen")
 RETURN g.geneSymbol`

Answer
`[[{"g.geneSymbol": "PPARG"}, {"g.geneSymbol": "BCL2"}]]`

You: what pathways are PPARG and BCL2 part of?

Query
`MATCH (g1:Gene {geneSymbol: 'PPARG'})-[:GENEINPATHWAY]-(:p:Pathway)
 MATCH (g2:Gene {geneSymbol: 'BCL2'})-[:GENEINPATHWAY]-(:p)
 RETURN p.pathwayName`

Answer
`[[{"p.pathwayName": "Pathways in cancer"}, {"p.pathwayName": "RXR and RAR heterodimerization with other nuclear receptor"}]]`

15

Querying AlzKB using KRAGEN

Launch AlzKB HOME ABOUT

The Alzheimer's KnowledgeBase (AlzKB)

Knowledge engineering is an important component of any AI project. This is especially true for the development and application of AI algorithms and software for the study of Alzheimer's disease where prior biological knowledge can assist with feature selection, model selection, and model interpretation. The Alzheimer's KnowledgeBase (AlzKB) integrates more than 20 different sources of knowledge about genes, pathways, drugs, and diseases to inform AI analyses. It is our hope that this integrated knowledgebase can be used to identify new biomarkers of Alzheimer's leading to new drug targets.

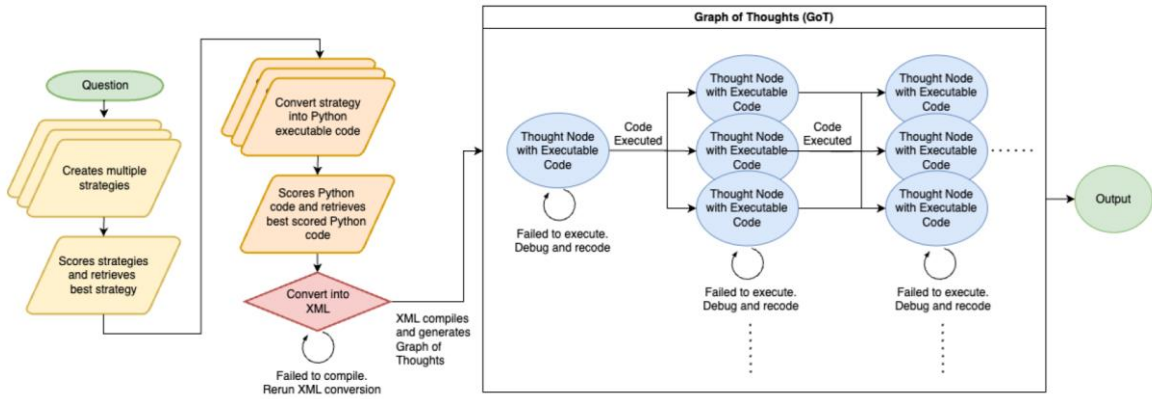
AlzKB was funded by grant R01 AG066833 from the National Institute on Aging (NIA), National Institutes of Health (NIH).

Cypher Query Search (Free)

Natural Language Search (OpenAI key required)

16

ESCARGOT: Dynamic Graph of Thoughts



ESCARGOT (in revision)

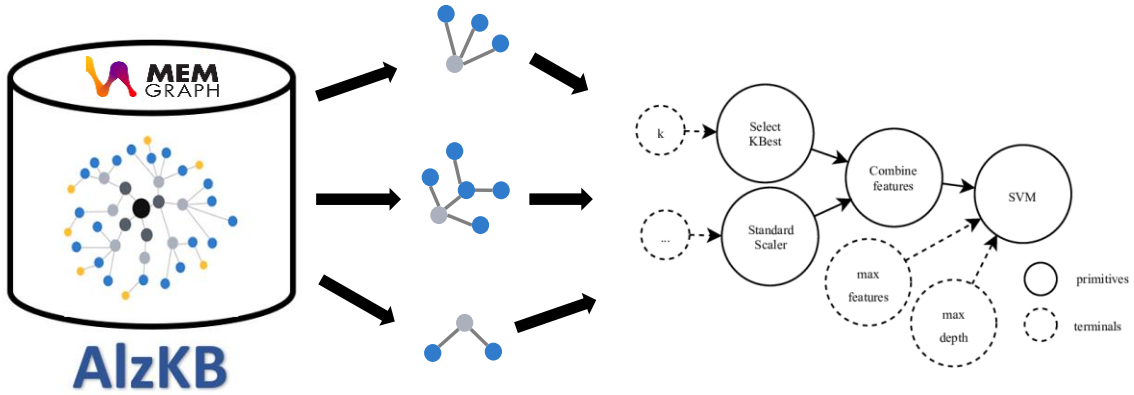
ESCARGOT: An AI Agent Leveraging Large Language Models, Dynamic Graph of Thoughts, and Biomedical Knowledge Graphs for Enhanced Reasoning

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,*}

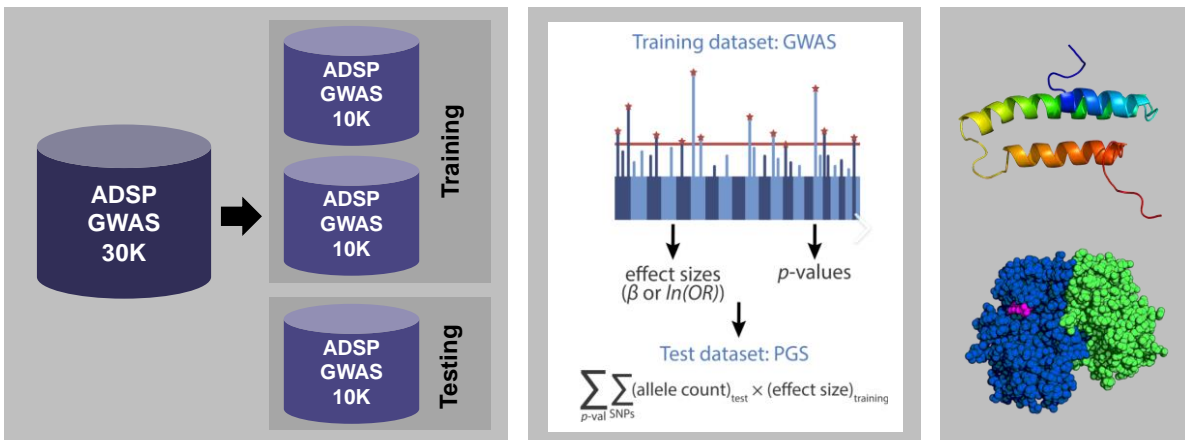
¹Department of Computational Biomedicine, Center for Artificial Intelligence Research and Education, Cedars Sinai Medical Center, 700 N. San Vicente Blvd., Pacific Design Center, Suite G-541H, West Hollywood, CA, USA

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%

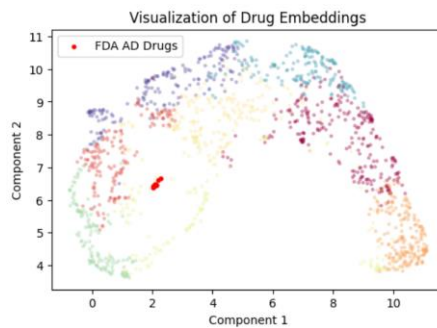
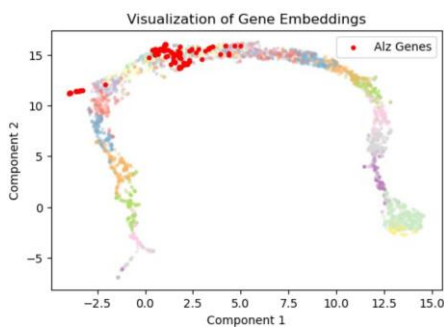
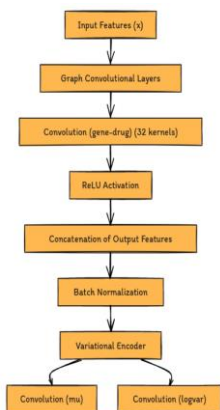
Automated Feature Selection and Machine Learning



Data Preparation (n=30k)

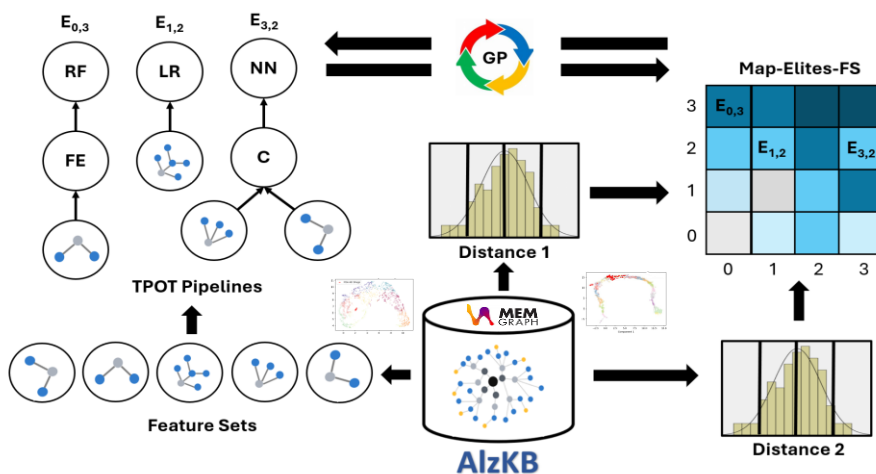


Defining Distance from Alzheimer's Using a Variational Graph Auto Encoder



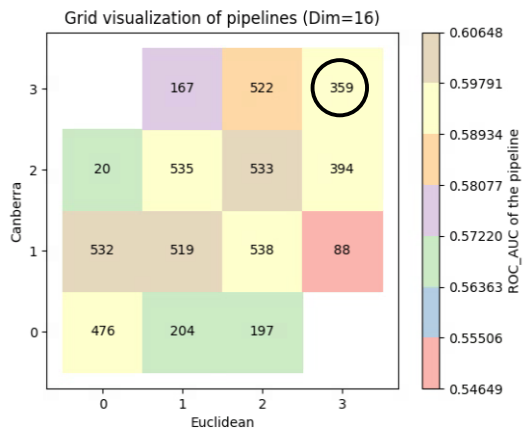
23

Knowledge-Guided Automated Machine Learning Using Map-Elites



24

Knowledge-Guided Automated Machine Learning Using Map-Elites



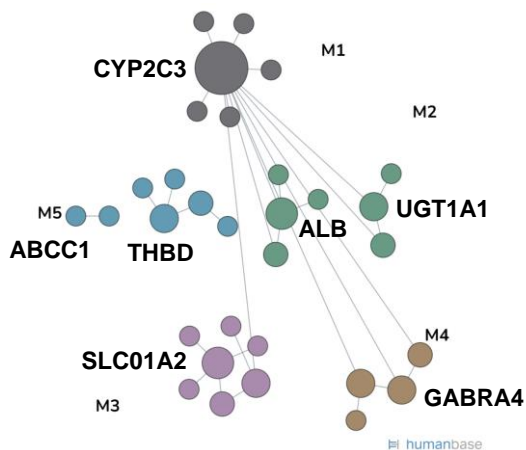
Feature set 359

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 SLC02B1
 UGT2B7 PTGS1 CYP2C9 CFTR CXCL8 PPARG
 SLC22A11 ABCB1 PLAT UGT1A1 SLC01A2 SLC22A8
 CYP2C19 UGT2B4 BCL2 UGT1A9 SLC22A6 PTGS2
 CYP2C8

Drugs

Temazepam (a benzodiazepine)
 Ibuprofen

Gene Interpretation Using HumanBase



M1: Drug Metabolism

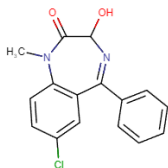
M2: Drug Metabolism

M3: Chloride Transport

M4: Chloride Transport

M5: Receptor Metabolism

Drug Interpretation



Temazepam, a benzodiazepine, targets the GABA neurotransmitter

NIH National Library of Medicine
National Center for Biotechnology Information

PubMed®

Temazepam AND alzheimer

Advanced Create alert Create RSS

Save Email Send to Sort by: Most recent Display options

MY NCBI FILTERS

RESULTS BY YEAR

TEXT AVAILABILITY

ARTICLE ATTRIBUTE

2 results

Page 1 of 1

The Incidence of Benzodiazepine and Related Drug Use in Persons with and without Alzheimer's Disease.
1
Cite: Saarela L, Taipale H, Koponen M, Tanskanen A, Tolppanen AM, Tiihonen J, Hartikainen S. J Alzheimer's Dis. 2016;49(3):809-18. doi: 10.3233/JAD-150630. PMID: 26484930
Share
BACKGROUND: Benzodiazepines and related drugs (BZDR) are occasionally used to treat certain symptoms of **Alzheimer's disease** (AD). However, the risks related to BZDR use are high in older persons. ...Incidence of BZDR, including benzodiazepines (lorazepam, oxa ...

From prescriptions to drug use periods - things to notice.
2
Cite: Tanskanen A, Taipale H, Koponen M, Tolppanen AM, Hartikainen S, Ahonen R, Tiihonen J. BMC Res Notes. 2014 Nov 14;7:796. doi: 10.1186/1756-0500-7-796. PMID: 25398553 **Free PMC article.**
Share
METHODS: We used 3,828,292 dispensed prescriptions claimed between 1 January 2002 and 31 December 2009 for 28,093 persons with **Alzheimer's disease**. Examples of drug use histories are presented to discuss different aspects that should be noticed when using reg ...
[View PDF](#)

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?

