Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

Grado en Máster Universitario en Innovación Digital

Complex Data in Health

Studying Alzheimer's Disease

Authors:

Emanuele Alberti
Leandro Duarte
Ottavia Biagi

December 2025

Contents

Introduction

For this assignment, we were assigned to study **Alzheimer's Disease (AD)**. The following identifiers were provided: UMLS CUI `C0002395`, NCIt code `C2866`, recommended protein Apolipoprotein E (UniProt: `P02649`), transcriptome dataset `GSE300079`, and image dataset from OASIS (Kaggle).

1 Medical Terminologies and Semantic Datasets

1.1 Disease General Information

We explored biomedical databases to gather standardized information about Alzheimer's Disease.

NCBI MedGen (UID: 1853) provides the following definition: *"A degenerative disease of the brain that causes dementia, which is a gradual loss of memory, judgment, and ability to function. This disorder usually appears in people older than age 65, but less common forms of the disease appear earlier in adulthood."* The disease is classified as *Disease or Syndrome* with concept ID `C0002395`. The directly associated gene is `APP` (21q21.3), with related genes including `APOE`, `PSEN1`, `PSEN2`, `ABCA7`, `MPO`, and `PLAU` [1].

NCBI MeSH (ID: D000544) describes AD as a degenerative brain disease with insidious onset of dementia, impairment of memory, judgment, and attention span, followed by apraxias and global loss of cognitive abilities. Pathologically, it is marked by senile plaques, neurofibrillary tangles, and neuropil threads [2].

ICD-10 classifies AD under code `G30`, with subtypes: `G30.0` (early onset, before age 65), `G30.1` (late onset, after age 65), `G30.8` (other), and `G30.9` (unspecified) [3].

Orphanet (ORPHA:1020) describes Early-onset autosomal dominant Alzheimer disease (EOAD), representing less than 1% of all AD cases, caused by mutations in `PSEN1` (69%), `APP` (13%), or `PSEN2` (2%) [4].

Table 1 summarizes the disease codifications across vocabularies.

Table 1: Alzheimer's Disease codifications across medical vocabularies [1, 2, 3, 5, 6, 7, 8, 4].

Vocabulary	Code
UMLS CUI	C0002395
MedGen UID	1853
MeSH	D000544
ICD-10	G30
SNOMED CT	26929004
NCIt	C2866
OMIM	104300, 516000
Orphanet	ORPHA:238616, ORPHA:1020
MONDO	MONDO:0004975
HPO	HP:0002511

1.2 SPARQL Query Results

We queried the NCIt SPARQL endpoint (`https://shared.semantics.cancer.gov/sparql`) to retrieve annotation properties for Alzheimer's Disease (code `C2866`) [6]. The query extracted the preferred label, synonyms, definition, semantic type, and UMLS CUI.

Summary of SPARQL Findings: The query retrieved annotation properties for NCIt code C2866. The disease is classified under semantic type "Mental or Behavioral Dysfunction" and maps to UMLS CUI C0002395. Seven synonyms were identified, including "Alzheimer's Dementia", "Alzheimer Disease", and "Alzheimer dementia". The definition describes AD as a progressive neurodegenerative disease characterized by nerve cell death leading to loss of cognitive function such as memory and language.

2 Bioinformatics

2.1 Disease Genes

We identified three main genetic factors linked to Alzheimer's disease from NCBI MedGen [1]: APP and PSEN1, when mutated, can directly cause rare, early-onset familial forms of the disease. APOE primarily affects an individual's risk for the more common late-onset form.

APP (Amyloid Precursor Protein) – NCBI Gene ID: 351, chromosome 21q21.3. The APP gene provides the blueprint for amyloid precursor protein. This protein is cut into smaller fragments, including amyloid-beta (Aβ), which is a key component of the plaques observed in the brains of Alzheimer's patients [9, 10]. Certain APP mutations or extra copies of the gene lead to an overproduction or more "sticky" forms of Aβ, which promotes plaque buildup and can directly cause early-onset familial Alzheimer's disease [9].

PSEN1 (Presenilin 1) – NCBI Gene ID: 5663, chromosome 14q24.2. PSEN1 encodes presenilin-1, which is the catalytic core of the γ-secretase enzyme. This enzyme performs the final cut of APP to release Aβ. Mutations in PSEN1 typically alter this cutting process, leading to the production of more of the longer, aggregation-prone Aβ peptides. This makes PSEN1 the most common genetic cause of autosomal-dominant early-onset Alzheimer's, with symptoms often beginning before age 65, and sometimes even before 40 [11, 12].

APOE (Apolipoprotein E) – NCBI Gene ID: 348, chromosome 19q13.32. APOE produces apolipoprotein E, a protein that assists in transporting fats and cholesterol within the body and brain [13]. Its common variants (ε2, ε3, ε4) differ in how they manage Aβ and brain lipid metabolism. Carrying the ε4 form increases the risk of late-onset Alzheimer's in a dose-dependent manner, while the ε2 variant tends to lower this risk. APOE primarily acts as a risk modifier; individuals with the ε4 allele may never develop Alzheimer's, and conversely, those without it can still develop the disease [13, 14].

Gene Sequence: Following the assignment recommendation, the APOE mRNA FASTA sequence (RefSeq: `NM_000041.4`) is provided in Appendix A [15].

A APOE Gene FASTA Sequence

>NM_000041.4 Homo sapiens apolipoprotein E (APOE), transcript variant 2, mRNA
CTACTCAGCCCCAGCGGAGGTGAAGGACGTCCTTCCCCAGGAGCCGACTGGCCAATCACAGGCAGGAAGA
TGAAGGTTCTGTGGGCTGCGTTGCTGGTCACATTCCTGGCAGGATGCCAGGCCAAGGTGGAGCAAGCGGT
GGAGACAGAGCCGGAGCCCGAGCTGCGCCAGCAGACCGAGTGGCAGAGCGGCCAGCGCTGGGAACTGGCA
CTGGGTCGCTTTTGGGATTACCTGCGCTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCA
GCTCCCAGGTCACCCAGGAACTGAGGGCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCCTACAAATC
GGAACTGGAGGAACAACTGACCCCGGTGGCGGAGGAGACGCGGGCACGGCTGTCCAAGGAGCTGCAGGCG
GCGCAGGCCCGGCTGGGCGCGGACATGGAGGACGTGTGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGC
AGGCCATGCTCGGCCAGAGCACCGAGGAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAA
GCGGCTCCTCCGCGATGCCGATGACCTGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGC
GCCGAGCGCGGCCTCAGCGCCATCCGCGAGCGCCTGGGGCCCCTGGTGGAACAGGGCCGCGTGCGGGCCG
CCACTGTGGGCTCCCTGGCCGGCCAGCCGCTACAGGAGCGGGCCCAGGCCTGGGGCGAGCGGCTGCGCGC
GCGGATGGAGGAGATGGGCAGCCGGACCCGCGACCGCCTGGACGAGGTGAAGGAGCAGGTGGCGGAGGTG
CGCGCCAAGCTGGAGGAGCAGGCCCAGCAGATACGCCTGCAGGCCGAGGCCTTCCAGGCCCGCCTCAAGA
GCTGGTTCGAGCCCCTGGTGGAAGACATGCAGCGCCAGTGGGCCGGGCTGGTGGAGAAGGTGCAGGCTGC
CGTGGGCACCAGCGCCGCCCCTGTGCCCAGCGACAATCACTGAACGCCGAAGCCTGCAGCCATGCGACCC
CACGCCACCCCGTGCCTCCTGCCTCCGCGCAGCCTGCAGCGGGAGACCCTGTCCCCGCCCCAGCCGTCCT
CCTGGGGTGGACCCTAGTTTAATAAAGATTCACCAAGTTTCACGCA

References

[1] NCBI. *MedGen UID 1853: Alzheimer's Disease*. https://www.ncbi.nlm.nih.gov/medgen/1853.

[2] NCBI. *MeSH D000544: Alzheimer Disease*. https://www.ncbi.nlm.nih.gov/mesh/68000544.

[3] WHO. *ICD-10 G30: Alzheimer's Disease*. https://icd.who.int/browse10/2019/en#G30.

[4] Orphanet. *Orphanet 1020: Early-onset autosomal dominant Alzheimer disease*. https://www.orpha.net/en/disease/detail/1020.

[5] NCBI. *SNOMED CT 26929004: Alzheimer disease*. https://vsac.nlm.nih.gov/context/cs/codesystem/SNOMEDCT/version/2021-09/code/26929004/info.

[6] NCI. *NCIt C2866: Alzheimer's Disease*. https://evsexplore.semantics.cancer.gov/evsexplore/concept/ncit/C2866.

[7] Monarch Initiative. *MONDO:0004975 Alzheimer disease*. https://monarchinitiative.org/MONDO:0004975.

[8] MSeqDR. *HPO HP:0002511 Dementia*. https://mseqdr.com/hpo_browser.php?2511.

[9] NCBI. *APP amyloid beta precursor protein [Homo sapiens (human)]*. https://www.ncbi.nlm.nih.gov/datasets/gene/351/. Gene ID: 351.

[10] Julia TCW and Alison M Goate. "Genetics of -Amyloid Precursor Protein in Alzheimer's Disease". In: *Cold Spring Harbor Perspectives in Medicine* (2017). DOI: 10.1101/cshperspect.a024539.

[11] NCBI. *PSEN1 presenilin 1 [Homo sapiens (human)]*. https://www.ncbi.nlm.nih.gov/datasets/gene/5663/. Gene ID: 5663.

[12] Jaya Bagaria, Eva Bagyinszky, and Seong Soo A An. "Genetics, Functions, and Clinical Impact of Presenilin-1 (PSEN1) Gene". In: *International Journal of Molecular Sciences* (2022). DOI: 10.3390/ijms231810970.

[13] NCBI. *APOE apolipoprotein E [Homo sapiens (human)]*. https://www.ncbi.nlm.nih.gov/datasets/gene/348/. Gene ID: 348.

[14] Amy C Raulin, Symone V Doss, Zachary A Trottier, Tadafumi C Ikezu, Guojun Bu, and Chia-Chen Liu. "APOE and Alzheimer's disease: advances in genetics, pathophysiology, and therapeutic approaches". In: *Molecular Neurodegeneration* (2022). DOI: 10.1186/s13024-022-00574-4.

[15] NCBI. *NM_000041.4 Homo sapiens apolipoprotein E (APOE), transcript variant 2, mRNA*. https://www.ncbi.nlm.nih.gov/nuccore/NM_000041.4?report=fasta.