DARG: A Database of Alzheimer Related Genes in Model Organisms

Julia Y. Song

Received April 01, 2025
Accepted July 07, 2025
Electronic access July 31, 2025

Alzheimer's disease affects a significant portion of the aging population worldwide. Despite extensive research over the years, a comprehensive understanding of its underlying mechanisms and effective treatments remain elusive. Meanwhile, animal models, such as mouse, *zebrafish*, *Drosophila*, and *C. elegans*, have proven invaluable in studying human diseases. In response, we have developed DARG, a Database of Alzheimer's disease-related genes in model organisms, designed to bridge the gap between human geneticists investigating the molecular mechanisms of the disease and the model organisms that can be used to explore the functions of disease-associated genes. DARG allows users to search and browse Alzheimer-related human genes from various resources and datasets, identify orthologs in model organisms, and access data on their gene expression in mouse and *Drosophila*, as well as the related phenotypes in *Drosophila*. This new resource will facilitate research projects to study the underlying molecular mechanisms of Alzheimer's disease using animal models.

Keywords: Alzheimer's disease, *Drosophila*, model organisms, database

Introduction

The research of Alzheimer's disease in animal models has significantly advanced our understanding of its mechanisms. The genetic manipulation to express human amyloid precursor protein (APP) and tau in various animal models, such as mouse, *zebrafish*, *Drosophila*, and *C. elegans*, causes Alzheimer-related features, including age-dependent amyloid-beta aggregation and neurodegeneration, making them powerful systems for studying disease pathogenesis[1–4]. For example, transgenic mouse models, such as the APP/PS1, 3xTg-AD, and tauopathies, are widely used to study key pathological features of Alzheimer's disease, including amyloid-beta plaque formation, tau tangles, neuroinflammation, and cognitive decline[5,6]. On the other hand, *Drosophila* and *C. elegans* models remain as valuable tools in Alzheimer's research with different advantages, including a simple nervous system, well-characterized genetics, and the ability to study aging and neurodegeneration in a short lifespan, despite their simplicity compared to vertebrate models. These models have also been used to explore genetic and environmental factors that contribute to Alzheimer's disease. For example, studies have shown that the manipulation of various genes, including those involved in insulin signaling and autophagy, can modulate amyloid-beta toxicity and improve cognitive function in *Drosophila* model[7,8].

In addition to studying disease mechanisms, animal models are also valuable for drug screening to identify potential therapeutic compounds that reduce amyloid-beta toxicity or enhance the clearance of amyloid plaques[9,10]. For example, curcumin has been shown to alleviate Alzheimer's disease by inhibiting inflammatory response, oxidative stress in mouse model, suggesting its potential as Alzheimer's treatments[11,12].

Informatics resources have been built in the past such as NIAGADS[13], AlzGene[14] and AlzBase[15] for users to search Alzheimer's related genes or data but the gene focused resources, such as AlzGene and AlzBase are currently not accessible anymore while NIAGADS is a data focused resource providing the data related to human genetics researches. To fill in the gap of gene focused resource as well as facilitate the research of Alzheimer's disease in model organisms, we built DARG (Database of Alzheimer Related Genes), an integrated resource collecting Alzheimer related human genes from various public resources, and mapped the human genes to the orthologous genes in model organisms. In addition, we also integrated the age-dependent and/or tissue specific transcriptomic data as well as phenotype annotation in nervous system from FlyBase and MGI (Mouse Genome Informatics). We believe this is a valuable resource for human geneticists investigating the disease's molecular mechanisms and researchers using the model organisms to explore the functions of disease-associated genes.

Result

Alzheimer's disease (AD) is a debilitating neurodegenerative disorder characterized by symptoms such as memory loss and cognitive decline. Its underlying mechanism involves a complex interplay of molecular and cellular processes, including amyloid-beta plaque formation, tau protein hyperphosphorylation, neuroinflammation, synaptic dysfunction, and oxidative stress. Some of the key molecular pathways involved in AD,

such as oxidative stress, are highly conserved across humans and various model organisms, which can be studied based on the orthologous genes directly in model organisms, whileas others are less conserved, such as Tau pathology, which can be studied expressing the exogenous human counterpart in transgenic animals (table 1). Different animal models offer distinct advantages (table 2) and have been instrumental in studying specific aspects of AD pathology. As a result, the research efforts using these models have significantly contributed to our understanding of disease mechanisms and the development of potential therapeutic strategies.

We collected Alzheimer associated genes from various public resources such as OMIM[16], GWAS[17], and ClinVar[18], and built DARG, a database for users to mine the list with ease (figure 1A). There are 2700 protein-coding genes in total collected (supplementary table 1), and we assigned confidence based on the number of resources as well as the publication counts if the gene is linked to two or more publications of Alzheimer focus. 388 (14%) genes from multiple resources are assigned high rank, while 720 (27%) are assigned moderate rank (figure 1B). Studying the function of Alzheimer's genes in animal models has proven to be important to advance our understanding of the molecular mechanisms of Alzheimer's disease[1–4]. To facilitate such studies, we also mapped the human Alzheimer related genes to their orthologs in the major model organisms using DIOPT[19]. Ortholog mapping is both complex and critical. DIOPT combined the results from around twenty ortholog prediction algorithms/resources, making it the most comprehensive tool of its kind, providing a more sensitive and specific mapping than any given resource could achieve. It used the number of tools that predict a given ortholog pair as the measurement of confidence. For example, using DIOPT mappings with high or moderate confidence, 2,670 (99%) of the human genes can be mapped to mouse orthologs, while 2,560 (95%) can be mapped to *zebrafish*. In comparison, 2,070 (77%) and 1,997 (74%) of the human genes can be mapped to *Drosophila* and *C. elegans*, respectively. As expected, a higher proportion of gene conservation is observed between species that are genetically closer to humans. Nonetheless, a substantial number of genes can still be studied across all of these model organisms (figure 1C), supporting their continued relevance in Alzheimer's disease research.

We examined the gene expression in *Drosophila* nervous system based on transcriptome datasets available at FlyBase[20], which were obtained from various samples of different developmental stages and/or genders, and observed that for 2047 human genes, the corresponding *Drosophila* ortholog is also expressing in nervous system in at least one of the samples. We included the average expression values of adult female and male flies from Day 1, 4, and 20 days respectively, making it possible to explore sex-biased expression in both young and aged flies at DARG. In addition, the mutant alleles of more than 50% of these *Drosophila* orthologs for 1072 human genes were also found to have abnormal neuroanatomy or neurophysiology phenotype, making *Drosophila* a great model to study the function of Alzheimer genes. For example, the human ERC2 gene was identified in two genome-wide association studies[21,22], but its role in Alzheimer's disease has not yet been investigated in either human or mouse models. *Drosophila* ortholog of human ERC2 is Brp (gene name: Bruchpilot), which is highly expressed in the adult nervous system and has been extensively studied in the context of neural function. Brp plays a critical role in regulating calcium channel clustering and synaptic vesicle release at the presynaptic active zone. Mutant alleles of Brp exhibit both neuroanatomical and neurophysiological phenotypes. Most importantly, in a *Drosophila* model of Alzheimer's disease, the expression of beta-amyloid was shown to cause an age-dependent reduction in Brp levels, shedding light on the mechanisms of synaptic impairment associated with beta-amyloid accumulation[23].

To evaluate the quality of the assembled list and understand the biological context underlying this gene list, we performed gene set enrichment analysis (GSEA) using PANGEA[24] with the high rank Alzheimer genes in DARG. The top gene sets from KEGG (Kyoto Encyclopedia of Genes and Genomes) annotation[25] that were highly over-represented included "Neurotrophin signaling pathway", "Apoptosis", "Lipid and atherosclerosis", "Alzheimer disease", while the top gene sets from gene group annotation of HGNC (HUGO gene nomenclature Committee)[26] enriched are "Dopamine receptors", "Neurotrophins", "Caspases", "Apolipoproteins", which are expected (figure 2). In addition, the top gene sets from the GO (gene ontology) biological process annotation included "aging", "protein maturation", "autophagy", while the top GO cellular component terms were "synapse", "cell junction", and "mitochondrion"[27] (figure 3). The GSEA results using the full list also showed similar results with less significant p values/fold changes (data not shown), which indicated that the functions of high rank genes in Alzheimer's are more extensively studied than moderate and low rank genes in the current research literature. The GSEA results were consistent with the existing knowledge of Alzheimer's disease, particularly regarding the associated biological processes, protein functions, and subcellular localizations.

While this alignment suggested that the assembled gene list captured relevant aspects of the disease and reflected current understanding, it did not by itself confirm the completeness or specificity of the gene set. Proteins and genes usually do not work alone, therefore, we also examined the protein complexes enriched among the high-ranking Alzheimer genes using COMPLEAT[28] (figure 4). For example, the protein complex HC4886 is identified in this analysis, which is a protein complex of ten members that positively regulates apoptosis, while HC8780, a protein complex of five members, regulates neuronal synaptic plasticity. The complex analysis results demonstrated the poten-

Table 1 AD Pathways Across Model Organisms

AD Pathways	Mouse	*zebrafish*	Fly	Worm
Amyloid-beta pathway	Conserved	Moderate conserved	Low conservation, use transgenic	Low conservation, use transgenic
Tau pathology	Conserved	Moderate conserved	Low conservation, use transgenic	Low conservation, use transgenic
Neuroinflammation	Conserved	Moderate conserved	Limited, simplified immune system (conserved innate immune but no adaptive immunity)	Limited, lack glial cells and canonical immune cells
Synaptic dysfunction	Conserved	Conserved	Conserved	Conserved
Oxidative stress/mitochondria	Conserved	Conserved	Conserved	Conserved

Table 2 Advantages and Disadvantages of Model Organisms

Advantages/disadvantages	Mouse	*zebrafish*	Fly	Worm
Conservation	high	moderate	low	low
Easy/difficult to work with	difficult	moderate	easy	easy
Fertility	low	high	high	high
Lifespan	long (1.5–3 year)	long (3–5 year)	short (40–60 day)	short (2–3 week)
Genetic modification	Moderate–Hard	Easy–Moderate	Very Easy	Extremely Easy
Time for genetic modification	months	weeks–months	2–4 weeks	1–2 weeks
Cost for genetic modification/maintain	high	high	low	low
Most important advantage	Present with AD-like pathology	Widely used in drug screen	Well annotated genome, established resources for genetic modification	Well-studied neuronal system

tial molecular mechanisms that underlie the related biological processes for Alzheimer's disease.

Discussion

The genes related to Alzheimer's disease from different research projects are scattered in many public resources, and there is an unmet need to integrate the candidate genes into a single comprehensive database. On the other hand, research on Alzheimer's disease in various model organisms has provided valuable insights into the molecular mechanisms underlying the disease and has facilitated the discovery of potential therapeutic targets. With this in mind, we developed an integrated database of human Alzheimer candidate genes with the information of orthologous genes in major model organisms. In addition, the transcriptomic datasets from mouse and *Drosophila* nervous systems, as well as the information on abnormal neuronal phenotype annotation in *Drosophila*, are integrated. Users can easily obtain a comprehensive list of candidate genes and identify the subset of genes that can be further studied in *Drosophila* or other model organisms.

Despite their advantages, animal models cannot fully replicate the complexity of Alzheimer's disease in humans. For example, all animals have anatomical and physiological differences from humans, lacking some of the Alzheimer's disease features observed only in humans, while *Drosophila* and *C. elegans* models have limitations due to their simpler nervous systems. However, ongoing research continues to refine these models and use them to better understand the molecular underpinnings of Alzheimer's disease, its genetic basis, and to screen for potential therapeutic agents. We believe DARG will bridge the gap between human geneticists studying the molecular mechanisms of the disease and researchers working with model organisms to explore the molecular functions of disease-related genes. Alzheimer's research is an ongoing effort, and as a result, the underlying source databases are periodically updated to reflect new findings. The DARG resource is scheduled for annual updates to incorporate these changes.

Methods

Retrieve human Alzheimer related genes from public resources

Alzheimer related genes were collected from public resources of human disease annotation including OMIM (An Online Catalog of Human Genes and Genetic Disorders, `https://omim.org/`[16], GWAS Catalog (An online catalog of human genome-wide association studies, `https://www.ebi.ac.uk/gwas/`[17], ClinVar (a public archive of reports of human variations classified for diseases and drug responses, with supporting evidence, `https://ftp.ncbi.nlm.nih.gov/pub/clinvar/gene_condition_source_id`[18], AGI (Alliance Genome Resources, `https://www.alliancegenome.org/`[29], UniProt `https://www.uniprot.org/`[30] and MalaCards `https://www.malacards.org/`[31] and BioLitMine `https://www.flyrnai.org/tools/biolitmine/web/`[32]. Information collection from various sources was performed in January 2025.

The collected information on associated genes was processed. Both reported genes and mapped genes were extracted from GWAS catalogs. Different resources might use different gene and protein identifiers; therefore, an ID mapping tool `https://www.flyrnai.org/tools/pangea/web/gene_map/9606` was used to map various identifiers to Entrez GeneID, and then integrated. The information about the number of publications co-citing both gene and Alzheimer's disease was retrieved from BioLitMine[32]. The genes supported by multiple resources were ranked "high" while the genes from a single resource with two or more publications with Alzheimer focus were assigned "moderate" rank. All the other genes were assigned "low" rank. We filtered out pseudo genes, ncRNAs, etc. to focus on protein-coding genes.

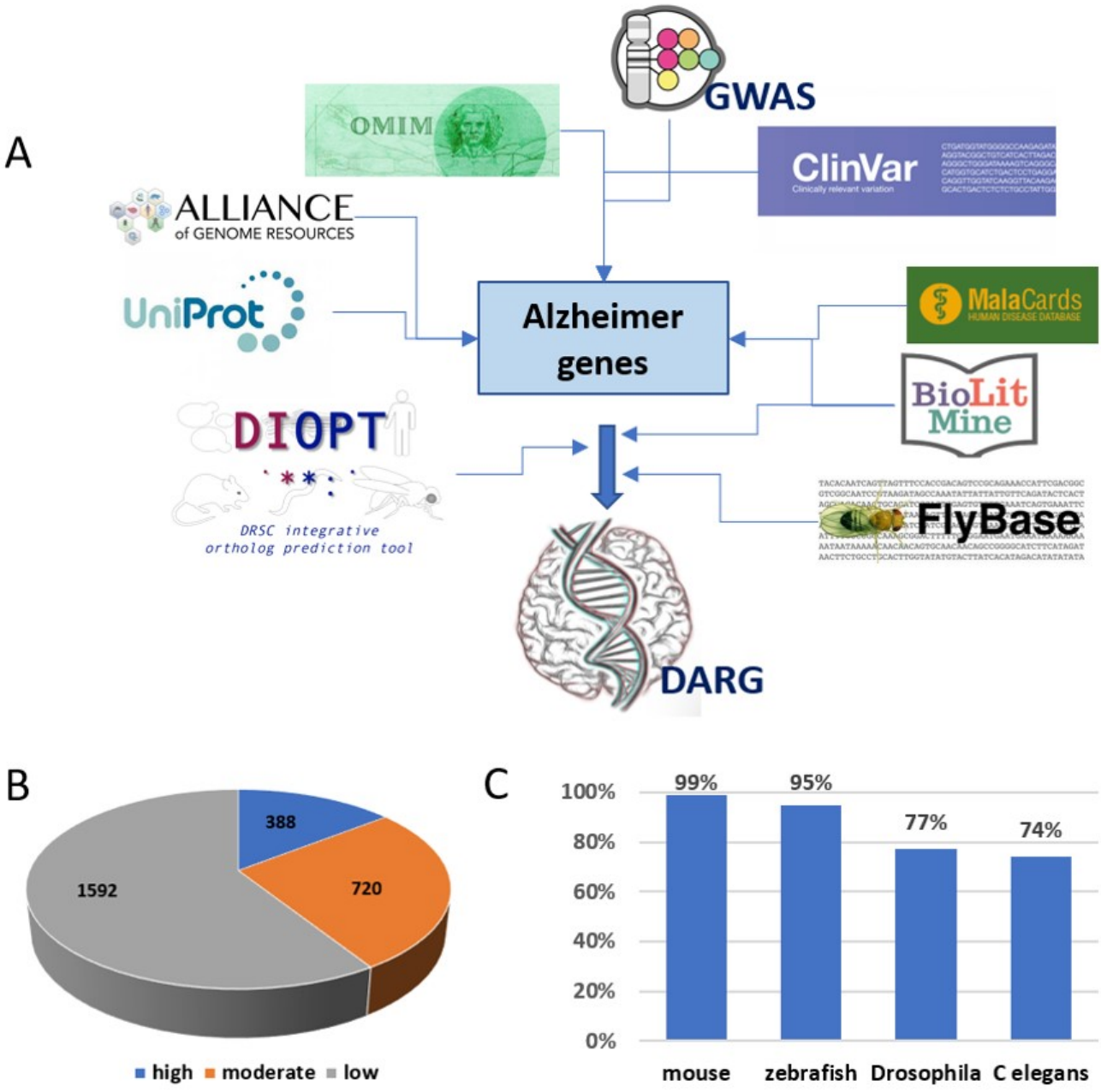

Fig. 1 Build the DARG database of Alzheimer genes. A.) workflow of database building and integrating the human genes related to Alzheimer's Disease from GWAS, OMIM and ClinVar supplemented with the genes annotated by MalaCards. Then the human genes were mapped to model organisms using DIOPT. B.) High rank is assigned to genes from multiple sources, while moderate rank genes are obtained from one resource with at least 2 papers co-cited with Alzheimer's Disease based on BioLitMine. C.) Percent of genes conserved in various model organisms based on DIOPT mapping with high or moderate confidence.

Map human Alzheimer related genes to model organisms

The assembled gene list was compared with orthologous relationships predicted by DIOPT[19], an integrated system for ortholog prediction with more than twenty algorithms integrated (eg. OrthoDB and Ensembl Compara), which uses the voting score as the measurement of mapping confidence. We only se-

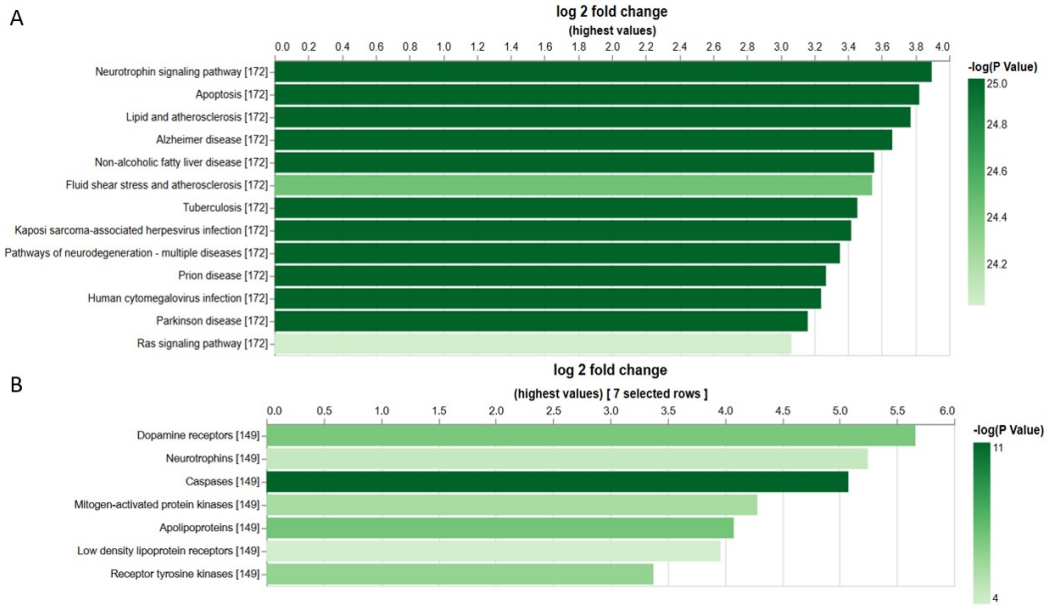

Fig. 2 Gene set enrichment analysis using PANGEA (https://www.flyrnai.org/tools/pangea/web/home/9606). Gene sets used are from A.) KEGG annotation B.) HGNC gene group annotation. The gene sets enriched were selected based on a cut-off of P value ¡0.05. The bar height reflects fold enrichment, while the darkness reflects enrichment P value.

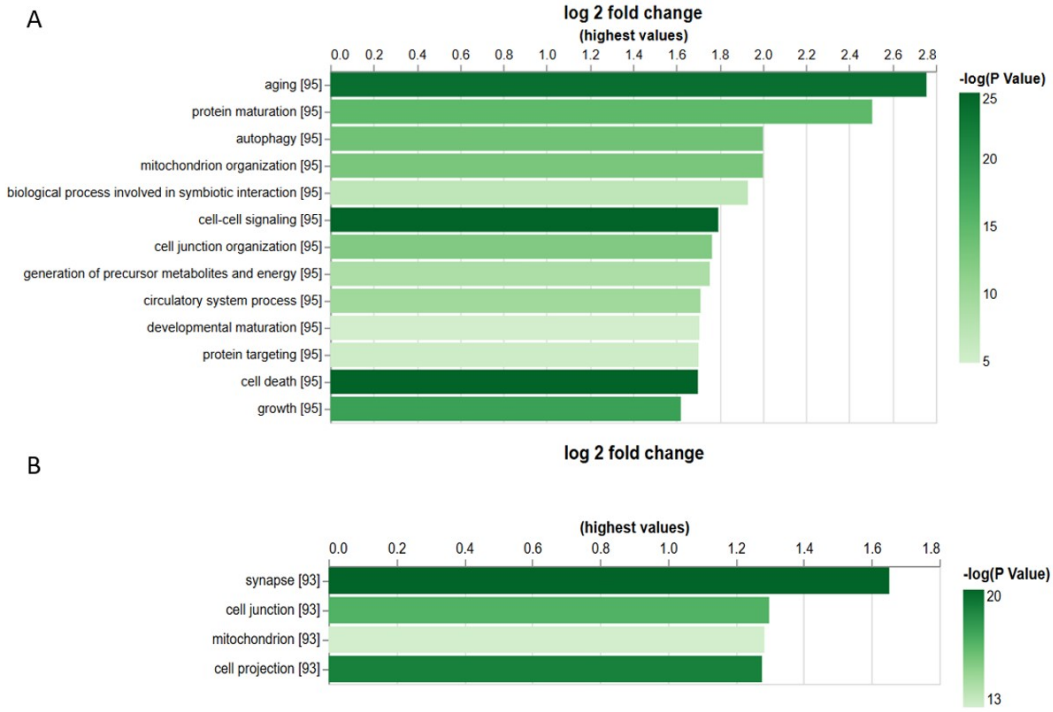

Fig. 3 Gene set enrichment analysis using PANGEA (https://www.flyrnai.org/tools/pangea/web/home/9606). Gene sets used are from A.) Biological process annotation of GO SLIM1 B.) Cellular component annotation of GO SLIM2. The gene sets enriched were selected based on a cut-off of P value <0.05. The bar height reflects fold enrichment, while the darkness reflects enrichment P value.

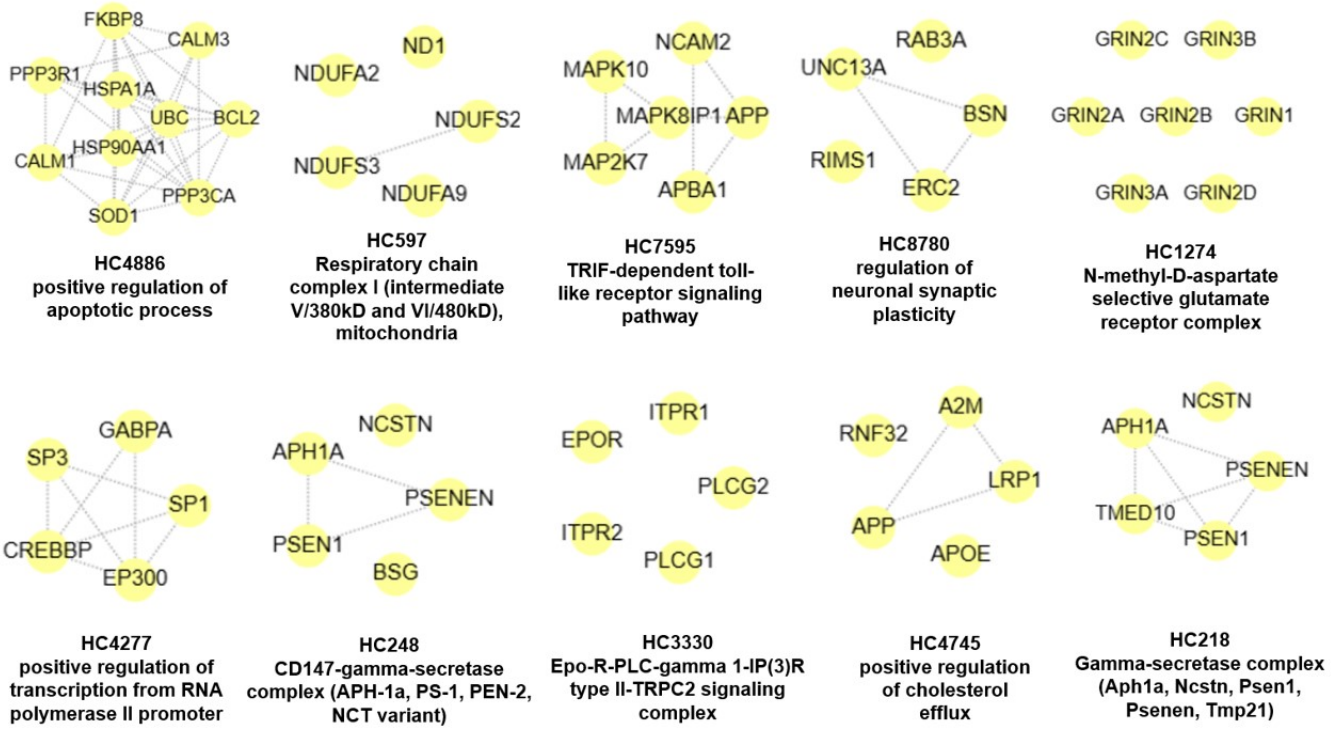

Fig. 4 Protein complex enrichment analysis using COMPLEAT `https://www.flyrnai.org/compleat/` : examples of over-represented protein complexes among high-ranking Alzheimer genes. The gene sets enriched were selected based on a cut-off of P value <0.05.

lected the predictions with high and moderate rank from DIOPT and mapped Alzheimer genes to mouse, *zebrafish*, *Drosophila*, and *C. elegans*. The orthologous genes with the highest DIOPT score from each of the model organisms are reported at the DARG site.

Retrieve tissue and stage specific transcriptomic data and phenotype annotation about orthologous genes in model organisms.

Tissue and stage specific RNA-seq datasets were obtained from the FlyBase ftp site (FB2024_06) `https://ftp.flybase.net/releases/FB2024_06/precomputed_files/`[20]. The datasets of the expression levels of the head samples from 1-day, 4-day, and 20-day adult flies were selected. The male and female datasets from day 1, 4, and 20 were averaged respectively, as well as combined and averaged at each time point. The genes involved in abnormal neuroanatomy or abnormal neurophysiology phenotype were retrieved from FlyBase (FB2024_06) `https://flybase.org/vocabularies` and integrated. The datasets of gene expression in the brain for adult mouse samples were retrieved from MGI (Mouse Gene Informatics) in May 2025 `https://www.informatics.jax.org/gxd`.

Gene set enrichment analysis (GSEA)

GSEA was performed using PANGEA[24] , and the gene set annotations selected were the KEGG pathway/disease annotation[25], HGNC gene group annotation[26], SLIM terms of biological process, and cellular component annotation of gene ontology[27]. Protein complex enrichment was done using COMPLEAT[28]. The statistical cutoff to select enriched gene sets was P value <0.05.

Database URL: `https://www.flyrnai.org/tools/AlzheimerGene`

Acknowledgments

I thank Dr. Claire Yanhui Hu at Harvard Medical School for the guidance and support, Mr. Eric Zhou for the help setting up the DARG web site. I also thank Harvard Medical School Research Computing for hosting the web site.

References

1 A. Alexander, V. Marfil and C. Li, *Use of Caenorhabditis elegans as a model to study Alzheimer's disease and other neurodegenerative diseases*.

2 G. Esquerda-Canals, L. Montoliu-Gaya, J. Guell-Bosch and S. Villegas, *Mouse Models of Alzheimer's Disease*.

3 B. Frost, J. Gotz and M. Feany, *Connecting the dots between tau dysfunction and neurodegeneration*.

4 B. Thawkar and G. Kaur, *The current models unravel the molecular mechanisms underlying the intricate pathophysiology of Alzheimer's disease using zebrafish*.

5 J. Jankowsky, T. Melnikova, D. Fadale, G. Xu, H. Slunt, V. Gonzales, L. Younkin, S. Younkin, D. Borchelt and A. Savonenko, *Environmental enrichment mitigates cognitive deficits in a mouse model of Alzheimer's disease*.

6 S. Oddo, A. Caccamo, J. Shepherd, M. Murphy, T. Golde, R. Kayed, R. Metherate, M. Mattson, Y. Akbari and F. LaFerla, *Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Abeta and synaptic dysfunction*.

7 Y. Huang, Z. Wan, Z. Wang and B. Zhou, *Insulin signaling in Drosophila melanogaster mediates Abeta toxicity*.

8 S. Oyarce-Pezoa, G. Rucatti, F. Munoz-Carvajal, N. Sanhueza, W. Gomez, S. Espinoza, M. Leiva, N. Garcia, D. Ponce, C. SanMartin, D. Rojas-Rivera, N. Salvadores, M. Behrens, U. Woehlbier, M. Calegaro-Nassif and M. Sanhueza, *The autophagy protein Def8 is altered in Alzheimer's disease and Abeta42-expressing Drosophila brains*.

9 P. Deshpande, N. Gogia and A. Singh, *Exploring the efficacy of natural products in alleviating Alzheimer's disease*.

10 D. Dam and P. Deyn, *Animal models in the drug discovery pipeline for Alzheimer's disease*.

11 S. Shao, X. Ye, W. Su and Y. Wang, *Curcumin alleviates Alzheimer's disease by inhibiting inflammatory response, oxidative stress and activating the AMPK pathway*.

12 Y. Zeng, Y. Lv, M. Hu, F. Guo and C. Zhang, *Curcumin-loaded hydroxypropyl-beta-cyclodextrin inclusion complex with enhanced dissolution and oral bioavailability for epilepsy treatment*.

13 A. Kuzma, O. Valladares, E. Greenfest-Allen, H. Nicaretta, M. Kirsch, Y. Ren, Z. Katanic, H. White, A. Wilk, L. Bass, J. Brettschneider, L. Carter, J. Cifello, W. Chuang, K. Clark, P. Gangadharan, J. Haut, P. Ho, W. Horng, T. Iqbal, Y. Jin, P. Keskinen, A. Rose, M. Moon, J. Manuel, L. Qu, F. Robbins, N. Saravanan, J. Sha, S. Tate, Y. Zhao, P. D. Sequencing, L. Cantwell and J. Gardner.

14 L. Bertram, M. McQueen, K. Mullin, D. Blacker and R. Tanzi, *Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database*.

15 Z. Bai, G. Han, B. Xie, J. Wang, F. Song, X. Peng and H. Lei, *AlzBase: an Integrative Database for Gene Dysregulation in Alzheimer's Disease*.

16 J. Amberger and A. Hamosh, *Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes*.

17 E. Sollis, A. Mosaku, A. Abid, A. Buniello, M. Cerezo, L. Gil, T. Groza, O. Gunes, P. Hall, J. Hayhurst, A. Ibrahim, Y. Ji, S. John, E. Lewis, J. MacArthur, A. McMahon, D. Osumi-Sutherland, K. Panoutsopoulou, Z. Pendlington, S. Ramachandran, R. Stefancsik, J. Stewart, P. Whetzel, R. Wilson, L. Hindorff, F. Cunningham, S. Lambert, M. Inouye, H. Parkinson and L. Harris, *The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource*.

18 M. Landrum, J. Lee, M. Benson, G. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, J. Hoover, W. Jang, K. Katz, M. Ovetsky, G. Riley, A. Sethi, R. Tully, R. Villamarin-Salomon, W. Rubinstein and D. Maglott, *ClinVar: public archive of interpretations of clinically relevant variants*.

19 Y. Hu, I. Flockhart, A. Vinayagam, C. Bergwitz, B. Berger, N. Perrimon and S. Mohr, *An integrative approach to ortholog prediction for disease-focused and other functional studies*.

20 J. Thurmond, J. Goodman, V. Strelets, H. Attrill, L. Gramates, S. Marygold, B. Matthews, G. Millburn, G. Antonazzo, V. Trovisco, T. Kaufman, B. Calvi and C. FlyBase, *FlyBase 2.0: the next generation*.

21 E. Adewuyi, E. O'Brien, D. Nyholt, T. Porter and S. Laws, *A large-scale genome-wide cross-trait analysis reveals shared genetic architecture between Alzheimer's disease and gastrointestinal tract disorders*.

22 C. Gouveia, E. Gibbons, N. Dehghani, J. Eapen, R. Guerreiro and J. Bras, *Genome-wide association of polygenic risk extremes for Alzheimer's disease in the UK Biobank*.

23 J. Huang, P. Ma, S. Ji, X. Zhao, J. Tan, X. Sun and F. Huang, *Age-dependent alterations in the presynaptic active zone in a Drosophila model of Alzheimer's disease*.

24 Y. Hu, A. Comjean, H. Attrill, G. Antonazzo, J. Thurmond, W. Chen, F. Li, T. Chao, S. Mohr, N. Brown and N. Perrimon, *PANGEA: a new gene set enrichment tool for Drosophila and common research organisms*.

25 M. Kanehisa, M. Furumichi, Y. Sato, M. Kawashima and M. Ishiguro-Watanabe, *KEGG for taxonomy-based analysis of pathways and genomes*.

26 E. Bruford, B. Braschi, L. Haim-Vilmovsky, T. Jones, R. Seal and S. Tweedie, *The importance of being the HGNC*.

27 C. Ontology, S. Aleksander, J. Balhoff, S. Carbon, J. Cherry, H. Drabkin, D. Ebert, M. Feuermann, P. Gaudet, N. Harris, D. Hill, R. Lee, H. Mi, S. Moxon, C. Mungall, A. Muruganugan, T. Mushayahama, P. Sternberg, P. Thomas, K. Auken, J. Ramsey, D. Siegele, R. Chisholm, P. Fey, M. Aspromonte, M. Nugnes, F. Quaglia, S. Tosatto, M. Giglio, S. Nadendla, G. Antonazzo, H. Attrill, G. Santos, S. Marygold, V. Strelets, C. Tabone, J. Thurmond, P. Zhou, S. Ahmed, P. Asanitthong, D. Buitrago, M. Erdol, M. Gage, M. Kadhum, K. Li, M. Long, A. Michalak, A. Pesala, A. Pritazahra, S. Saverimuttu, R. Su, K. Thurlow, R. Lovering, C. Logie, S. Oliferenko, J. Blake, K. Christie, L. Corbani, M. Dolan, H. Drabkin, D. Hill, L. Ni, D. Sitnikov, C. Smith, A. Cuzick, J. Seager, L. Cooper, J. Elser, P. Jaiswal, P. Gupta, P. Jaiswal, S. Naithani, M. Lera-Ramirez, K. Rutherford, V. Wood, J. Pons, M. Dwinell, G. Hayman, M. Kaldunski, A. Kwitek, S. Laulederkind, M. Tutaj, M. Vedi, S. Wang, P. D'Eustachio, L. Aimo, K. Axelsen, A. Bridge, N. Hyka-Nouspikel, A. Morgat, S. Aleksander, J. Cherry, S. Engel, K. Karra, S. Miyasato, R. Nash, M. Skrzypek, S. Weng, E. Wong, E. Bakker, T. Berardini, L. Reiser, A. Auchincloss, K. Axelsen, G. Argoud-Puy, M. Blatter, E. Boutet, L. Breuza, A. Bridge, C. Casals-Casas, E. Coudert, A. Estreicher, M. Famiglietti, M. Feuermann, A. Gos, N. Gruaz-Gumowski, C. Hulo, N. Hyka-Nouspikel, F. Jungo, P. Mercier, D. Lieberherr, P. Masson, A. Morgat, I. Pedruzzi, L. Pourcel, S. Poux, C. Rivoire, S. Sundaram, A. Bateman, E. Bowler-Barnett, A. Bye, P. Denny, A. Ignatchenko, R. Ishtiaq, A. Lock, A. Ponferrada, S. Ramachandran, L. Ruzicka and M. Westerfield, *The Gene Ontology knowledgebase in 2023*, S. Diamantakis, D. Raciti, M. Zarowiecki, M. Fisher, C. James-Zorn, V.

28 A. Vinayagam, Y. Hu, M. Kulkarni, C. Roesel, R. Sopko, S. Mohr and N. Perrimon, *Protein complex-based analysis framework for high-throughput data sets*.

29 C., *Alliance of Genome Resources. Updates to the Alliance of Genome Resources central infrastructure*.

30 C. UniProt, *UniProt: the Universal Protein Knowledgebase in 2023*.

31 N. Rappaport, M. Twik, I. Plaschkes, R. Nudel, T. Stein, J. Levitt, M. Gershoni, C. Morrey, M. Safran and D. Lancet, *MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search*.

32 Y. Hu, V. Chung, A. Comjean, J. Rodiger, F. Nipun, N. Perrimon and S. Mohr, *BioLitMine: Advanced Mining of Biomedical and Biological Literature About Human Genes and Genes from Major Model Organisms*.