Structural analysis on mutations related to Alzheimer's disease

Antigoni Avramouli, Eleftheria Polychronidou, Panayiotis Vlamos✉

BiHELab – Bioinformatics and Human Electrophysiology Lab, Department of Informatics of Ionian University, Corfu, Greece
Competing interests: AA none; EP none; PV none

Abstract

Proteins have a significant role in all biological processes. The functional properties of proteins rely upon their three-dimensional structures. Over the last twenty years substantial advances in genomic technologies have enhanced our knowledge of the genetics of Alzheimer's disease. To that end, the identification of mutations pathogenicity is still of vital importance. The methodology of the present research work focuses on the structural analysis of proteins related to Alzheimer's disease and the comparative study to create groups with clear structural similarity and pathogenicity. To achieve that, three-dimensional descriptors (fpfh, rsd and 3dsc) were applied along with supervised machine learning classification methods. In total, 62 APP, 286 PSEN1, 68 PSEN2 and 25 MAPT variants were evaluated in our study. The output of the methodology characterised thirty mutations that were unclear at the point of the data collection.

Introduction

Alzheimer's disease (AD), the most common neurodegenerative disease, is identified by an insidious decline in cognitive and memory function. Furthermore, the number of AD is growing rapidly with the increase in the aging population (Saez-Atienzar *et al.*, 2020). AD has a long prodromal phase, thus the onset of the pathogenetic changes until the appearance of clinical symptoms makes early diagnosis and treatment of this disease more demanding. The diagnosis criteria to define AD based on biomarker evidence currently include deposits of extracellular senile plaques in the cerebral cortex, the formation of neurofibrillary tangles, and neurodegeneration [AT(N)] classification system. As a result, there is still on-going research for improving the identification and classification of AD patients. To that end, the current research approach aims to deliver a methodology that can predict the classification of AD through the pathogenicity of the mutation. The pathogenicity is further mapped to clinical phenotypes that can support the patients' stratification and the prediction of disease progression. The current methodology was applied to four proteins: APP, MAPT, PSEN1 and PSEN2. Three of them are associated with autosomal-dominant AD, amyloid precursor protein (APP) (OMIM 104760), presenilin 1 (PSEN1) (OMIM 104311) and presenilin 2 (PSEN2) (OMIM 600759), while microtubule associated protein tau (MAPT) (OMIM

157140) encodes the tau protein that is aberrantly phosphorylated in AD (Neuner *et al.*, 2020).

Background

Based on the age of onset, AD is divided into two classes: early-onset AD (EOAD) with onset before 65, and late-onset AD (LOAD) (Cuyvers and Sleegers, 2016); EOAD comprises about 5% to 10% of all AD patients and has strong patterns of familial inheritance (Zhu *et al.*, 2015). EOAD has been linked to pathogenic mutations in one of three causative genes: APP, PSEN1, PSEN2. Up to now, over 400 known mutations on these genes have been described, while PSEN1 mutations are responsible for approximately 75% of genotyped families positive for a mutation, whereas APP and PSEN2 mutations account for 13% and 12%, respectively. Aβ peptides result from the cleavage of APP by β- and γ-secretases while PSEN1 and PSEN2 are components of the γ-secretase complex (Haass and De Strooper, 1999).

APP encodes for the amyloid precursor protein, a transmembrane protein whose cleavage forms amyloidogenic Aβ peptides, key components of amyloid plaque. Most APP mutations are missense or nonsense. They are normally localised either within the domain that encodes the Aβ peptide, (amino acids 692–705) (93% of total mutations) or near the cleavage sites of secretases (amino acids 670–682 and 713–724) (Cacace *et al.*, 2016; Dai *et al.*, 2017). The overall effect of APP

Research Papers

mutations alters the processing by secretases and leads to increased generation and/or aggregation of amyloid, and/or a change in the ratio of specific Aβ peptides.

Presenilin-1 and presenilin-2 proteins are critical subunits of the γ-secretase complex responsible for processing of APP. PSEN2 is about 60% homologous to PSEN1, thus it is possible that they also have overlapping or similar activities. Similar to some mutations in APP, mutations in PSEN1 and PSEN2 typically result either in the overproduction of Aβ or an increased ratio of Aβ42 over Aβ40 (Loy *et al.*, 2014), triggering the formation of amyloid plaques and leading to the development of AD (Sun *et al.*, 2017). Mutations in PSEN1 are the most common cause of EOAD; as of September 2021, over 350 mutations (some of unclear pathogenicity) have been identified (www.Alzforum.org). PSEN1 mutations are estimated to contribute to around 80% of monogenic AD with complete penetrance and early age of onset (Giri *et al.*, 2016). The exact mechanism through which mutations in PSEN1 result in dementia and neurodegeneration in EOAD remains unknown. In addition to their role in γ-secretase activity, PSEN1 mutations may compromise neuronal function, affecting γ secretase activity and kinesin-I-based motility, thus leading to neurodegeneration (Giri *et al.*, 2016). To date, 341 pathogenic mutations have been identified in PSEN1, most of whom are missense, while most of them occur in exons 5, 6, 7, and 8.

PSEN2 mutations are much rarer, with only around 30 mutations identified in EOAD families (Cacase *et al.*, 2016). Mutations in PSEN2 alter the γ-secretase activity and lead to elevation of Aβ42/40 ratio in a similar manner to the PSEN1 mutation. Though PSEN2 is homologous to PSEN1, less amyloid peptide is produced by PSEN2 mutations. In some people with PSEN2 mutations, neuropathological changes appear as neuritic plaque formation and neurofibrillary tangle accumulation (Giri *et al.*, 2016). Furthermore, β-secretase activity is enhanced by PSEN2 mutation, through reactive oxygen species-dependent activation of extracellular signal-regulated kinase (Park *et al.*, 2012). PSEN2 mutations are very rare, and to date 84 pathogenic PSEN2 mutations have been detected worldwide. Moreover, in the pathogenic/likely pathogenic variants, missense variants are more common in PSEN1 than those in PSEN2. In general, most of the pathogenic AD mutations are located in exons 16–17 of the APP, exons 3–12 of PSEN1 and PSEN2 genes (An *et al.*, 2016). The localisation of mutations in AD causing genes leads to the assumption that the above exons are variant hotspots and need to be given priority when performing DNA sequencing (Zhao and Liu, 2017).

MAPT encodes the microtubule associated protein tau, a protein crucial to AD neuropathology. Even though MAPT mutations are not linked to familial forms of AD, SNPs near the MAPT locus are associated with AD risk. Interestingly SNPs in exon 3 act protectively against AD through decreased aggregation of tau protein (Neuner *et*

al., 2020). Up to now 15 no disease-causative mutations have been linked with AD.

Methodology

The implementation methodology for this research work follows the approach of the Automated shape-based clustering of 3D immunoglobulin protein structures that was evaluated in the use case of chronic lymphocytic leukemia (Polychronidou *et al.*, 2018).

The proteins described were used as target proteins due to their important role in Alzheimer's disease progression. The first step of the analysis was to identify the mutations related to the proteins. This information was extracted by the Alzforum - Mutations public database (Alzforum, 2021). This database is a repository of variants in genes linked to Alzheimer's disease (AD). The database includes the three genes associated with autosomal-dominant AD (APP, PSEN1, PSEN2) and two genes associated with AD by way of genetics or the neuropathology of the encoded protein (TREM2 and MAPT). TREM2 was excluded from the analysis as the identification of the primary structure wasn't feasible by the selected sources.

Evaluation of Protein Structures

Since the three-dimensional shape of most of the related proteins is not determined through experimental methodologies, established servers and online databases like Uniprot (UniProt Consortium, 2015), PolyPhen-2 (Adzhubei *et al.*, 2013), iTASSER (Yang *et al.*, 2015) and PDBeFold (Krissinel, 2007) were evaluated for predicting the mutated structures and estimate the impact of the mutations to the 3-dimensional structure. A list of the selected methodologies is presented on the Table 1

These methodologies were used to better understand the structures and determine the structures for the mutated proteins. However, during the study, AlphaFold (AlQuraishi, 2019) was published as the latest state-of-the-art method for the prediction of protein structures. AlphaFold (Pereira, 2021), a neural network-based model, was validated in the challenging 14th Critical Assessment of Protein Structure Prediction (CASP14) and was vastly more accurate than competing methods. The four protein structures were identified in the AlphaFold database and used as the target structures of this analysis.

The following step in the pipeline was to create the structures for the mutations by protein. To achieve that the DynaMut server was used (Rodrigues *et al.*, 2018). DynaMut implements two distinct, well established normal mode approaches, which can be used to analyse and visualise protein dynamics by sampling conformations and assess the impact of mutations on protein dynamics and stability resulting from vibrational entropy changes. DynaMut integrates our graph-based signatures along with normal mode dynamics to generate

a consensus prediction of the impact of a mutation on protein stability.

Through this approach the mutated structures were predicted for the four proteins of interest. Specifically, 25 structures were retrieved by MART mutations, 62 structures by APP mutations, 286 structures by PSEN1 mutations and 68 structures by PSEN2 mutations. For each mutation, the description on Pathogenicity was also extracted from ALZforum and normalised into four categories: (1) Unclear, (2) Benign, (3) Pathogenic, (4) Not classified.

The objective of the study was to classify the resulted structures by Unclear mutations, through an AI/ML approach derived by the protein structures. To further analyse the mutated structures, an established methodology from the field of 3D object recognition was applied. The individual examination and combination of the local descriptors was applied to the 3D structures to extract the appropriate features for the comparison.

Three distance matrices were created by applying the FPFH, RSD and 3DSC descriptors. These matrices were the input of hierarchical clustering. The methodology was selected to supervise the separation of structures into clusters of structures with high similarity.

Indicative examples of hierarchical clustering output are presented here for the APP structure (Figure 1), PSEN2 structure (Figure 2), and the 3DSC descriptor. The optimal number of clusters for each protein-descriptor combination was determined through Silhouette analysis (Figure 3) and the clusters were analyzed based
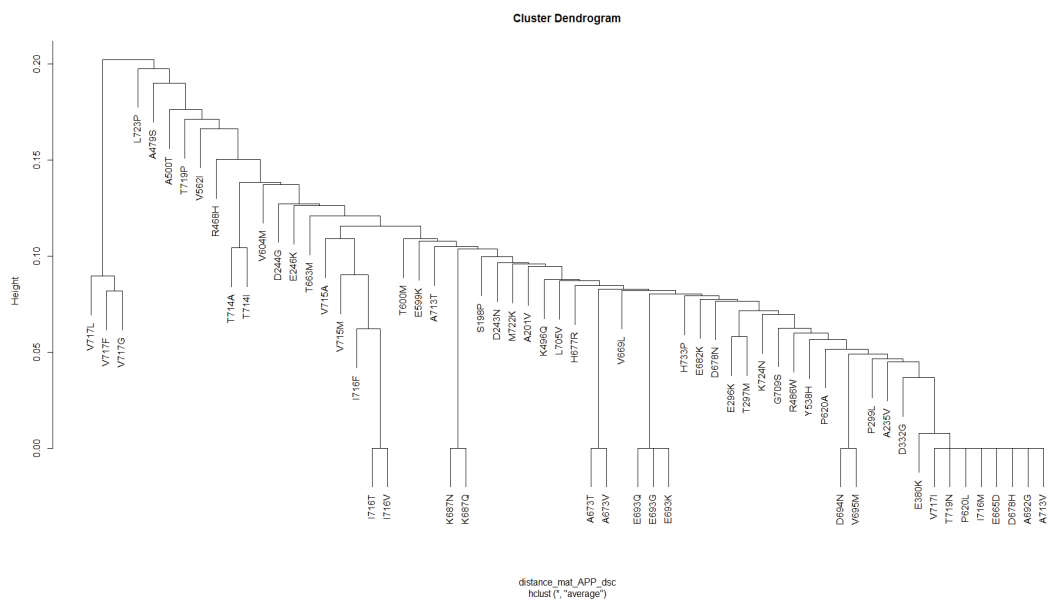

Figure 1. Dendrogram resulted from the hierarchical clustering in APP protein using the 3DSC descriptor.
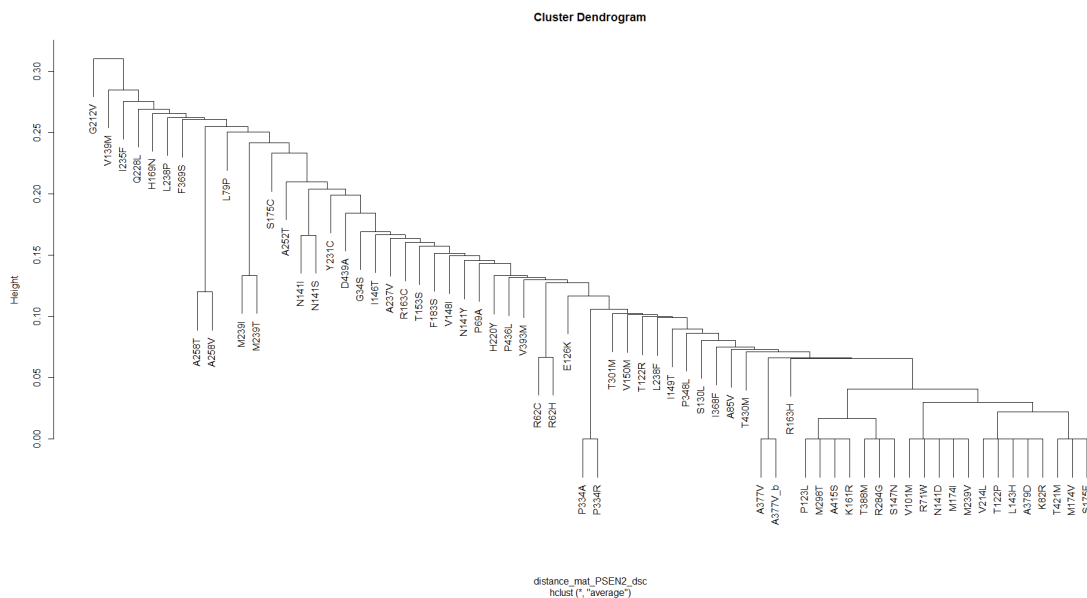

Figure 2. Dendrogram resulted from the hierarchical clustering in PSEN2 protein using the 3DSC descriptor.
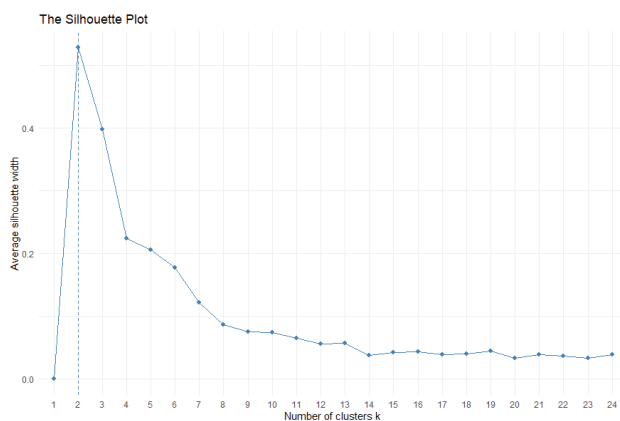
Research Papers

Figure 3. Silhouette analysis on PSEN1 protein using the FPFH descriptor, to determine the optimal number of clusters.
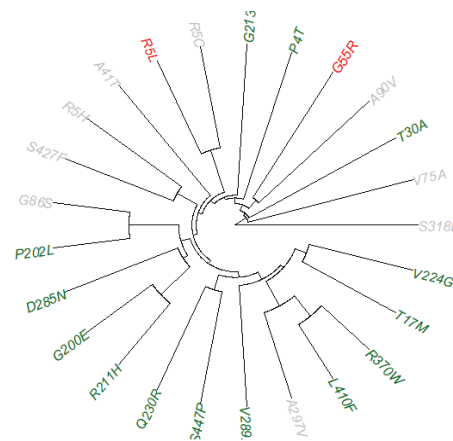

Figure 4. Fan dendrogram of MAPT protein using the 3DSC descriptors.

on their pathogenicity. The results from this analysis classified the PSEN1 mutations into two main clusters, Pathogenic and Non-pathogenic. Based on this analysis, 110 mutated structures originally derived by Unclear pathogenic mutations, were classified as pathogenic.

To further analyse the output, fan dendrograms were produced by also using the colors of the pathogenic types (Figure 4). Through this low-level cluster visualisation, the lowest height of the cluster was identified, and the groups of protein structures were analysed. In detail, six unclear structures were characterised for APP, nine unclear structures were characterised for PSEN2 and four unclear structures were characterised for MAPT.

3DSC descriptor supported the characterisation of the nine unclear or not classified mutations related to

PSEN2. On MAPT, two mutations were characterised as Pathogenic (A90V, R5C) by RSD and 3DSC while R5C was classified as Benign from FPFH. R5H was classified as pathogenic by RSD and unclear by FPFH. Finally, A297V was classified as Benign by all methods and G86S as Benign only by 3DSC as the other descriptors didn't reveal any specific cluster. In MAPT case FPFH and RSD didn't perform as 3DSC in the cases of established pathogenicity (unclear cases), thus 3DSC is the descriptor that performed best in this protein. By following the output of the 3DSC descriptor, four new mutations can be characterised.

In the case of APP, FPFH was the descriptor with the highest confidence and through this approach the method characterised six Not Classified mutations. This

Table 1. Results of PSEN2 pathogenicity prediction. The numbers on the descriptors column describes the groups that the proteins were grouped with while the color describes the type of pathogenicity (green = Benign, red=Pathogenic). The new prediction of the not classified or unclear structures is included in the corresponding cell.

Original Protein	Mutation	Pathogenicity	3DSC
PSEN2	A379D	Not Classified -> Pathogenic	4
PSEN2	A415S	Not Classified -> Pathogenic	1
PSEN2	K161R	Not Classified-> Pathogenic	1
PSEN2	K82R	Not Classified-> Pathogenic	4
PSEN2	L143H	Not Classified-> Pathogenic	4
PSEN2	M174I	Not Classified-> Pathogenic	3
PSEN2	M174V	Benign	5
PSEN2	M239V	Likely Pathogenic	3
PSEN2	M298T	Uncertain Significance	1
PSEN2	N141D	Not Classified-> Pathogenic	3
PSEN2	P123L	Likely Pathogenic	1
PSEN2	S175F	Uncertain Significance	5
PSEN2	T122P	Likely Pathogenic	4
PSEN2	T421M	Benign	5
PSEN2	V101M	Unclear Pathogenicity -> Pathogenic	3
PSEN2	V214L	Unclear Pathogenicity-> Pathogenic	4

Table 2. Results of APP pathogenicity prediction

Original Protein	Mutation	Pathogenicity	FPFH
APP	A235V	Likely Benign	5
APP	A692G	Pathogenic	6
APP	E296K	Not Classified -> Benign	5
APP	E380K	Uncertain Significance	7
APP	E665D	Benign	1
APP	G709S	Not Classified	9
APP	H733P	Not Classified -> Pathogenic	6
APP	I716M	Not Classified	9
APP	K496Q	Not Classified-> Benign	5
APP	L705V	Pathogenic	8
APP	M722K	Pathogenic	2
APP	P299L	Not Classified -> Pathogenic	2
APP	P620L	Uncertain Significance	7
APP	R486W	Not Classified-> Pathogenic	3
APP	T297M	Uncertain Significance	8
APP	T719N	Pathogenic	4
APP	V562I	Uncertain Significance	9
APP	V669L	Not Classified-> Benign	1
APP	V717F	Pathogenic	3
APP	V717I	Pathogenic	4
APP	V717L	Pathogenic	8

number corresponds to ~20% of the total not classified APP mutations.

To support the analysis of the methodology, evidence for clinical phenotype, pathogenicity, neuropathology, and biological effect were also taken into consideration. For example, in APP - H733P mutation has not been classified, but the in-silico analysis suggests damaging effect (Guerreiro *et al.*, 2010). This mutated structure was classified as pathogenic by this process as well. Hence, additional evidence beyond the experimental evaluation is generated by our suggested methodology.

Discussion

In summary, the phenotype of APP, PSEN1 and PSEN2 mutation carriers is heterogeneous. Applying pathogenicity prediction methodology to variants of unknown significance, we classified many of them as probably pathogenic. Variants of unknown significance were mainly identified in single individuals' phenotype clinically with AD. Data from families with a monogenic form of AD or patients with a known causative mutation provide the opportunity to identify mutation-specific effects and to correlate genotypic changes with clinical and pathophysiological manifestations of the disease. Asymptomatic carriers of mutations can also serve as candidates for disease-modifying treatment or prevention trials. In the future, different genetic causes of AD should be targeted with specific interventions.

Studies involving mapping pathogenic mutations to tertiary structural domains are required to show the vital relationships between structure and function. Since the amino acid position can, in fact, predict pathogenicity we analysed mutations in AD causative genes and compared these changes to available clinical data. To the best of our knowledge, this is the first study of its kind performing comparative and ab initio prediction of protein structure for mutated APP, PSEN1, PSEN2 and MAPT proteins. In this study we used prediction tools to elucidate how mutations in the causative genes change the tertiary structure of the proteins. We aim in the identification of common structural issues, and in the relation between structure and function through the deleterious effects of the loss of tertiary structure in EOAD causative genes.

Key Points

- There is still on-going research for improving the identification and classification of Alzheimer's disease patients.
- Structural similarity of mutated proteins supports the evidence generation for characterisation of mutations pathogenicity.
- The applied implementation uses three-dimensional descriptors to identify the distance between the structures.
- The methodology was very effective and successfully generated a new dimension in the pathogenicity determination process.

Avramouli et al. (2022) EMBnet.journal 27, e1011

http://dx.doi.org/10.14806/ej.27.0.1011

References

Adzhubei I, Jordan DM, Sunyaev SR (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics, **76**(1), 7-20. http://dx.doi.org/10.1002/0471142905.hg0720s76

ALZFORUM (2021) Mutations https://www.alzforum.org/mutations Accessed 20 Sep 2021

AlQuraishi M (2019) AlphaFold at CASP13. Bioinformatics, **35**(22), 4862-4865. http://dx.doi.org/10.1093/bioinformatics/btz422

An SS, Park SA, Bagyinszky E, Bae SO, Kim YJ, Im JY *et al.* (2016) A genetic screen of the mutations in the Korean patients with early-onset Alzheimer's disease. Clin Interv Aging **15**, 1817-1822. http://dx.doi.org/10.2147/CIA.S116724

Cacace R, Sleegers K, Van Broeckhoven C (2016) Molecular genetics of early-onset Alzheimer's disease revisited. Alzheimers Dement **12**(6), 733–748. http://dx.doi.org/10.1016/j.jalz.2016.01.012

Cuyvers E and Sleegers K (2016) Genetic variations underlying Alzheimer's disease: evidence from genome-wide association studies and beyond. Lancet Neurol **15**(8), 857–868. http://dx.doi.org/10.1016/S1474-4422(16)00127-7

Dai MH, Zheng H, Zeng LD, Zhang Y (2017) The genes associated with early-onset Alzheimer's disease. Oncotarget **9**(19), 15132–15143. http://dx.doi.org/10.18632/oncotarget.23738

Giri M, Zhang M, Lü Y (2016) Genes associated with Alzheimer's disease: an overview and current status. Clin Interv Aging **11**, 665–681. http://dx.doi.org/10.2147/CIA.S105769

Guerreiro RJ, Baquero M, Blesa R, Boada M, Brás JM *et al.* (2010) Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP. Neurobiol aging **31**(5), 725–731. http://dx.doi.org/10.1016/j.neurobiolaging.2008.06.012

Haass C and De Strooper B (1999) The presenilins in Alzheimer's disease--proteolysis holds the key. Science **286**(5441), 916–919. http://dx.doi.org/10.1126/science.286.5441.916

Krissinel E (2007). On the relationship between sequence and structure similarities in proteomics. Bioinformatics **23**, 717-723. http://dx.doi.org/10.1093/bioinformatics/btm006

Loy CT, Schofield PR, Turner AM, Kwok JB (2014) Genetics of dementia. Lancet **383**(9919), 828–840. http://dx.doi.org/10.1016/S0140-6736(13)60630-3

Neuner SM, Tcw J, Goate AM (2020) Genetic architecture of Alzheimer's disease. Neurobiol Dis **143**, 104976. http://dx.doi.org/10.1016/j.nbd.2020.104976

Park MH, Choi DY, Jin HW, Yoo HS, Han JY *et al.* (2012) Mutant presenilin 2 increases β-secretase activity through reactive oxygen species-dependent activation of extracellular signal-regulated kinase. J Neuropathol Exp Neurol **71**(2), 130-139. http://dx.doi.org/10.1097/NEN.0b013e3182432967

Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM *et al.* (2021). High-accuracy protein structure prediction in CASP14. Proteins: Structure, Function, and Bioinformatics, **89**(12), 1687-1699. http://dx.doi.org/10.1002/prot.26171

Polychronidou E, Kalamaras I, Agathangelidis A, Sutton LA, Yan XJ *et al.* (2018) Automated shape-based clustering of 3D immunoglobulin protein structures in chronic lymphocytic leukemia. BMC Bioinformatics **19**(14), 67-81. http://dx.doi.org/10.1186/s12859-018-2381-1

Rodrigues CH, Pires DE, Ascher DB (2018) DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. Nucleic Acids Res **46**(W1), W350-W355. http://dx.doi.org/10.1093/nar/gky300

Saez-Atienzar S and Masliah E (2020) Cellular senescence and Alzheimer disease: the egg and the chicken scenario. Nat Rev Neurosci **21**(8), 433–444. http://dx.doi.org/10.1038/s41583-020-0325-z

Sun L, Zhou R, Yang G, Shi Y (2017) Analysis of 138 pathogenic mutations in presenilin-1 on the in vitro production of Aβ42 and Aβ40 peptides by γ-secretase Proc Natl Acad Sci U S A **114**(4), E476–E485. http://dx.doi.org/10.1073/pnas.1618657114

UniProt Consortium (2015) UniProt: a hub for protein information. Nucleic acids research **43**(D1), D204-D212. http://dx.doi.org/10.1093/nar/gku989

Yang J, Yan R, Roy A, Xu D, Poisson J *et al.* (2015). The I-TASSER Suite: protein structure and function prediction. Nature methods **12**(1), 7-8. http://dx.doi.org/10.1038/nmeth.3213

Zhao GH and Liu XM (2017) Clinical features and genotype-phenotype correlation analysis in patients with ATL1 variants: A literature reanalysis. Transl Neurodegener **6**:9. http://dx.doi.org/10.1186/s40035-017-0079-3

Zhu XC, Tan L, Wang HF, Jiang T, Cao L *et al.* (2015) Rate of early onset Alzheimer's disease: a systematic review and meta-analysis. Ann Transl Med **3**(3), 38. http://dx.doi.org/10.3978/j.issn.2305-5839.2015.01.19

Research Papers