Association of early-onset Alzheimer’s disease with germline-generated high affinity self-antigen load

Poyin Huang1,2,3,4, Yuan-Han Yang3,5, Ya-Hsuan Chang6, Shu-Ling Chang7, Mei-Chuan Chou5, Chiou-Lian Lai1, Ching-Kuan Liu1 and Hsuan-Yu Chen5,8

Abstract
Self-antigen presentation outside the central nervous system has crucial role regarding self-proteins tolerance and autoimmunity, leading to neuroinflammation. Self-antigen with strong-binding affinity is considered to be pathogenic. We aim to investigate whether strong-binding affinity self-antigen load is associated with early/late-onset Alzheimer’s disease (AD). A total of 54 AD samples (22 early-onset, 32 late-onset) underwent next-generation sequencing (NGS) for whole-exome sequencing. Genotypes of HLA class I genes and germline mutations were obtained for estimation of the binding affinity and number of self-antigens. For each patient, self-antigen load was estimated by adding up the number of self-antigens with strong-binding affinity. Self-antigen load of early-onset AD was significantly higher than late-onset AD (mean ± SD: 6115 ± 2430 vs 4373 ± 2492; p = 0.011). An appropriate cutoff value 2503 for dichotomizing self-antigen load was obtained by receiver operating characteristic (ROC) curve analysis. Patients were then dichotomized into high or low self-antigen load groups in the binary multivariate logistic regression analysis. Adjusted odds ratio of the high self-antigen load (>2503) was 14.22 (95% CI, 1.22–165.70; p = 0.034) after controlling other covariates including gender, education, ApoE status, and baseline CDR score. This is the first study using NGS to investigate germline mutations generated self-antigen load in AD. As strong-binding affinity self-antigen is considered to be pathogenic in neuroinflammation, our finding indicated that self-antigen load did have a role in the pathogenesis of AD owing to its association with neuroinflammation. This finding may also contribute to further research regarding disease mechanism and development of novel biomarkers or treatment.

Introduction

Alzheimer’s disease (AD) is the most common cause of cognitive decline of the elderly. AD is pathologically defined by the accumulation of extracellular β-amyloid plaques and intracellular neurofibrillary tangles in the brain along with extensive neuronal loss. However, increasing evidence also shows that AD pathogenesis is not only limited to β-amyloid plaques and intracellular neurofibrillary tangles, but also includes strong immunological interactions within the brain.

The accumulation of β-amyloid plaques and toxic tau protein could activate microglia and induce innate immune responses characterized by release of inflammatory mediators, which further contributes to AD progression.

In human genome, the human leukocyte antigen (HLA) gene contains the most polymorphic region, as it codes for antigens directly involved in immunological recognition. The high polymorphism of HLA gene is responsible for the complicated ability of the immune system to react to a tremendous variety of different peptides. As immune responses are based on the capability of HLA molecules to differentially bind specific peptides, the strength of immune responses differs according to the presentation of
variable HLA antigens. Thus, specific HLA molecules induce distinct immune responses and may have a role in regulating susceptibility to certain diseases such as AD, supported by some recent studies showing that some specific HLA alleles are involved in the pathogenesis of AD. Furthermore, striking evidences indicate that the immune system is intrinsically involved in the pathogenesis of AD. Auto-antibodies against a variety of molecules have been identified in AD patients, but the roles of these auto-antibodies in the pathogenesis of AD are yet unclear.

Many concepts regarding central nervous system (CNS) immune function have been challenged and changed in recent years, including but not limited to the discoveries of CNS-draining lymphatic vessels, the origin and functional diversities of microglia, the influence of T cells on CNS homeostasis to immunology, and the role neuro-inflammation plays in neurodegenerative diseases. On the other hand, self-antigen presentation outside the CNS may have crucial role regarding self-proteins tolerance and autoimmunity leading to neuroinflammation. Based on the above theories, we hypothesize that germline mutation generated self-antigens will have a negative role in AD, to the opposite of somatic mutation generated antigens, which play a positive role in cancer treatment. The aim of this study is to utilize cutting edge next-generation sequencing (NGS) technique, the whole-exome sequencing, to generate HLA-based self-antigen load in AD patients and test whether it has a role in AD.

Materials and methods

Study population and DNA extraction

A total of 54 AD patients were collected from the neurological outpatient department of Kaohsiung Medical University Hospital. The diagnosis of AD was based on the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) criteria. The patients with AD had to fulfill the following criteria in order to be included in the probable group in this study: (1) a diagnosis of dementia established by clinical and neuropsychological examinations; (2) progressive cognitive impairments present in two or more areas of cognition without conscious disturbance; (3) onset after the age of 40 and before the age of 90 years; and (4) brain CT or MRI studies compatible with AD. The included patients were taking donepezil 5–10 mg/d, rivastigmine 3–6 mg/12 hours, galantamine 8–12 mg/12 hours, or memantine 5–10 mg/12 hours. An annual interview was performed with each patient, and the Clinical Dementia Rating (CDR) score (0 (normal) to 3 (severe stage)) was used to evaluate their functional deterioration. The patients with an increased CDR score during follow-up were defined as having clinical progression. The clinical characteristics including age, gender, education level, ApoE genotype, serial CDR score, baseline stages of AD, and medications used for AD were recorded. Genomic DNA was extracted from the buffy coat following standard protocols.

This study was approved by the Institutional Review Board of our hospital, and written informed consent was obtained from all patients or their relatives before participating in this study. The above method has been described in detail in our previous publication.

Whole-exome sequencing analysis and HLA genotyping

The NGS platform HiSeq 4000 (Illumina, San Diego, CA) was performed for whole-exome sequencing analysis with SureSelect Human All Exon V6 kit (Agilent Technologies, Santa Clara, CA). The adaptor sequences were first removed from the raw sequencing reads data. In the quality control for all reads, lower-quality bases (sliding window with four-base length) were further trimmed from the raw sequence reads using software Trimomatic (v0.33). Trimmed reads were mapped to hg19 human reference genome using Burrows-Wheeler Aligner (v0.7.5a). In order to reduce false positive results, realignment, quality score recalculation and marking duplication possible from PCR were done by GATK (v3.7) and Picard (v1.98). After the quality controls, removed duplications, and, mapping results, the average reads throughput was 100× depth. The HaplotypeCaller method of GATK was used to detect germline mutation. Based on the germline mutations of each sample, the Polysolver software was carried out to estimate genotypes of HLA class I genes-A, B, and C, respectively. This algorithm extracted reads in HLA region and aligned them to the database of HLA full-length genomic sequences in the international immunogenetics information system (http://www.imgt.org). Based on the read alignment status, insert size distribution, base qualities, and information of Asian population allele frequency distribution, each sample’s class I HLA genotypes were estimated.

HLA genotyping of Taiwanese general population

In order to obtain HLA-A, HLA-B, and HLA-C genotypes of Taiwanese general population for comparison with AD patients, genotypes of HLA class I genes were retrieved from Taiwan Biobank (https://www.twbiobank.org.tw/new_web_en/index.php).

A total of 1075 individuals with HLA genotype data were available for HLA-A, 1090 individuals were available for HLA-B and 1089 individuals were available for HLA-C. These HLA genotypes were then used to compare with HLA genotypes of AD patients.

Prediction of the binding affinity of self-antigen

Based on the detected germline mutations, ANNOVAR software was used to annotate identified variants. After
the above analysis, germline mutations that could cause peptide change (nonsynonymous mutations) were further retrieved. After the annotation step, each nonsynonymous mutation was annotated how amino-acid changing and the mutant position of the protein (e.g., YAP1 R331W was amino acid R to W changed in the position 331 on protein YAP1). Because the binding peptide length of HLA class 1 was nearly 8–14 amino acids, a total of 13 amino acids on upper and downstream of the mutant site were selected. Hence, the length of the selected peptide was 27 amino acids (e.g., for YAP1 R331W, selected peptide was from position 318–344 and centered on 331). Finally, non-synonymous mutations derived peptide sequences and HLA genotypes were further input to NetMHC 4.0 Server (http://www.cbs.dtu.dk/services/NetMHC/) to predict binding affinity of self-antigen for AD patients that could be presented by HLA (Fig. 1). The number of the peptides with strong-binding affinity was the self-antigen load and the unit of the self-antigen load was peptide number.

**Statistical analysis**

The Fisher’s exact test or Mann–Whitney U test was used to compare differences between groups for categorical or continuous variables, respectively. Receiver operating characteristic (ROC) curve was used to find an appropriate cutoff value for self-antigen load. Binary logistic regression was applied for multivariate analysis using early/late-onset as dependent variable while self-antigen load, gender, years of education, baseline CDR score, and ApoE genotype were added to the model as covariates. For the power calculation, given the sample sizes for early and late-onset AD = 22 and 32, type 1 error = 0.02, and crude odds ratio = 12.6, the power of this study is 0.78. It is close to 0.8, which is the commonly used power value. All tests were two-sided and a p value < 0.05 was considered significant.

**Results**

A total of 22 early-onset and 32 late-onset AD patients were recruited in this study for analysis. All patients were Han Chinese and no relationship between samples. The basic characteristics of these 54 patients were shown in Table 1. In the early-onset group, average education years were borderline higher (8.88 ± 3.83 vs 6.38 ± 4.95; p = 0.084) and baseline dementia stage of the diagnosis time was more severe (lower percentage of CDR 0.5–1, 68.2% vs 96.9%; p = 0.006). In addition, early-onset group had more patients with high self-antigen load (95.5%) than late-onset group (62.5%) (p = 0.008). There were no significant differences between gender and the status of ApoE ε4+. However, self-antigen load of early-onset AD was significantly higher than that of late-onset (average ±
SD: 6115 ± 2430 vs 4373 ± 2492; p = 0.011) (Table 1 and Fig. 2).

We also obtained HLA class I genotypes of Taiwanese general population (1075 cases for HLA-A, 1090 cases for HLA-B, and 1089 cases for HLA-C) from Taiwan Bio-bank. Alleles of HLA-A, HLA-B, and HLA-C were compared between AD patients and general population. Results showed that HLA-A*11:01 (28.7% and 25.2%), HLA-B*40:01 (21.3% and 18.8%) and HLA-C*07:02 (33.3% and 19.4%) (Table 2) were the most major HLA class I genotypes for both AD patients and general population. When using HLA genotype as a binominal variate (for example, HLA-A*02:01 vs non HLA-A*02:01) for further analysis between early and late-onset AD (categorical, early vs late), we found that HLA-A*02:01 (early-onset 31.8% vs late-onset 6.3%; p = 0.023) was identified to be associated with the age of AD onset.

ROC curve was further performed to find that self-antigen load 2503 was an appropriate cutoff value for binary multivariate logistic regression analysis (Supplementary Fig. 1). Patients were dichotomized into high or low self-antigen load groups in the binary multivariate logistic regression analysis (Table 3). Adjusted odds ratio of the high self-antigen load (>2503) was 14.22 (95% CI, 1.22–165.70; p = 0.034) after controlling other covariates including gender, education, ApoE status, and baseline CDR score.

**Discussion**

In the past one to two decades of AD research regarding amyloid plaques have been fraught with disappointment. Several years focusing on amyloid plaques, the hallmark of AD ultimately resulted in no significant progress toward therapeutic strategies and prevention. Recently, pharmaceutical companies have announced that their trials of drugs meant to help block the production of amyloid plaques had
Some researchers have turned their focus to neuroinflammation as the main target because it is killing the bulk of the neurons that leads to AD\textsuperscript{15}. However, neuroinflammation through MHC I molecules, could negatively regulate cognitive function in the hippocampus with an age-dependent pattern. Furthermore, systemic \(\beta\)-2-microglobulin accumulating in aging blood may promote age-associated cognitive function decline and impair neurogenesis through MHC I\textsuperscript{16}. Wang et al. identified HLA gene variants including TNF-\(\alpha\) rs2534672, rs2395488, HFE rs1800562, and RAGE rs2070600 might involve in the structural change of brain regions that are associated with AD, thus modulating the susceptibility of AD\textsuperscript{7}. In this study, we identified the HLA genotypes of AD patients and compared with general population within the same ethnic groups. Although HLA class I genotype distribution is significantly different between AD and general population, HLA-A*11:01, HLA-B*40:01, and HLA-C*07:02 are the most major genotypes for both groups. We also identified that HLA-A*02:01 was associated with the age of AD onset, but as HLA genotypes are too diverse and the sample size is small, this finding is only for reference and HLA genotypes could not be used as a practical biomarker. The major contribution of this study is that we found that self-antigen load was significantly associated with age of AD onset. This result is compatible with the hypothesis that germline mutation generated self-antigens did have a negative role in AD, as early-onset patients have significantly higher self-antigen load.

Auto-antibodies could be produced under both physiological and pathophysiological situations. Under normal physiological situation, humans will generate natural auto-antibodies to recognize self-antigens in order to promote the identification and clearance of dead and dying cells\textsuperscript{20}. However, under pathophysiological situation, high affinity auto-antibodies will be produced. The mechanism of high affinity auto-antibodies production is mainly through inflammatory responses that activate affinity maturation of antibody through targeting self-antigens\textsuperscript{21}. Through binding to high affinity self-antigens, pathogenic auto-antibodies launch and retain the immuno-inflammatory cascade responsible for tissue damages. Though the CNS is previously considered to be an immune privileged site, recent research has shown that a defined lymphatic and glympathic system is capable of draining self-antigens and initiate CNS-directed adaptive immunological reactions\textsuperscript{15}. The trigger of T-cell mediated adaptive immunological reactions against self-antigens depends mainly on antigens drainage to structures like the lymph nodes of deep cervical chain, the leptomeningeal compartments and the choroid plexus\textsuperscript{22}. Therefore, T cells are capable of infiltrating the CNS and influence immune homeostasis in a pathogenic manner\textsuperscript{23}. When immunological disturbance is presented in the CNS, T cells may become pathogenic when inflammation persisted\textsuperscript{24}. As to the age of AD onset, only 4–5\% AD patients develop before age 65, the so called early-onset AD\textsuperscript{25}. At present, the cause of early-onset AD is not clear. Mutations of several genes including amyloid protein precursor (APP), presenilin-1 (PSEN1), and presenilin-2 (PSEN2) are thought to be responsible for autosomal dominant forms of early-onset AD. All these genes are involved in generation of \(\beta\)-amyloid\textsuperscript{26}. However, most
early-onset AD is sporadic and only around 4–5% of patients with early-onset AD have an autosomal dominant mutation in APP, PSEN1, and PSEN2 gene. Recent research indicates that early-onset AD is quite a heterogeneous disorder\(^a\). Based on our finding, it is reasonable to state that besides those well known genetic mutations involved in generation of β-amyloid, more-severe neuroinflammation may also contribute to the earlier onset of AD, as the self-antigen load estimated in this study is with strong-binding affinity, which is considered to be pathogenic.

To our knowledge, this is the first study based on NGS whole-exome sequencing to study germline mutation derived self-antigen load in AD. Our finding indicated that self-antigen load did have a role in the pathogenesis of AD probably owing to its association with neuroinflammation. This finding may also contribute to further research, regarding disease mechanism and development of novel biomarkers or treatment.

Acknowledgements
The authors would like to thank the Integrated Core Facility for Functional Genomics and Pharmacogenomics Lab, and Mathematics in Biology Group of Institute of Statistical Science AS, and National Center for High-performance Computing of National Applied Research Laboratories in Taiwan (MOST108-2319-B-492-001) supported genomic, biological, bioinformatics work, and computational and storage resources. This study is supported by grants from Kaohsiung Medical University, Kaohsiung Medical University Research Center Grant (KKMU-TC108B01), Ministry of Science and Technology, Taiwan, 108-2320-B-037-006- Kaohsiung Municipal Siaogang Hospital, Kmhk-104-041.

Author details
1Department of Neurology, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan. 2Department of Neurology, Kaohsiung Medical University, Kaohsiung, Taiwan. 3Department of Neurology, Faculty of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan. 4Department of Neurology, Kaohsiung Municipal Tung-Ho Hospital, Kaohsiung, Taiwan. 5Institute of Statistical Science, Academia Sinica, Taipei, Taiwan. 6School of Post-Baccalauraeate Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan. 7Ph.D. Program in Microbial Genomics, National Chung Hsing University, Taichung, Taiwan.

Author contributions
Poyin Huang, Yuan-Han Yang, Ya-Hsuan Chang, Shu-Ling Chang, Mei-Chuan, Chiou-Lan Lai, Ching-Kuan Liu, and Hsuan-Yu Chen contributed to the interpretation of the data/critical revision of the manuscript. All authors have no significant financial interest/other relationship to disclose.

Data availability
Whole-exome sequencing data and de-identification clinical data were available on http://ad.stat.sinica.edu.tw.

Conflict of interest
The authors declare that they have no conflict of interest.

Publisher’s note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at [https://doi.org/10.1038/s41398-020-0826-6].

References
1. Karch, C. M & Goate, A. M. Alzheimer’s disease risk genes and mechanisms of disease pathogenesis. Biol Psychiatry 77, 43–51 (2015).
2. Heneka, M. T. et al. Neuroinflammation in Alzheimer’s disease. Lancet Neurol. 14, 388–405 (2015).
3. Rosenberg, R. N. et al. Genomics of Alzheimer disease: a review. JAMA Neurol. 73, 867–874 (2016).
4. Raghavan, N. & Tosto, G. Genetics of Alzheimer’s disease: the importance of polygenic and epistatic components. Curr. Neurol. Neurosci. Rev. 17, 78 (2017).
5. Smith, L. K. et al. β2-microglobulin is a systemic pro-aging factor that impairs cognitive function and neurogenesis. Nat. Med. 21, 932–937 (2015).
6. Wang, Z. X. et al. Genetic association of HLA gene variants with MRI brain structure in Alzheimer’s disease. Mol. Neurobiol. 54, 3195–3204 (2017).
7. Steele, N. Z. et al. Fine-mapping of the human leukocyte antigen locus as a risk factor for Alzheimer disease: a case-control study. PLoS Med. 14, e1002272 (2017).
8. Lu, R. C. et al. Association of HLA-DRB1 polymorphism with Alzheimer’s disease: a replication and meta-analysis. Oncotarget 8, 93219–93226 (2017).
9. Padmadas, N., Panda, P. K. & Durairaj, S. Binding patterns associated All-HSP60 p458 conjugate to HLA-DRβ1 allele of human in Alzheimer’s disease: an in silico approach. Intersci. Sci. 10, 93–104 (2018).
10. Schettini, S. T. T., Gomez-Nicolà, D., Garcia-Vallejo, J. J. & Van Kooij, Y. Neur inflammation: microglia and T cells get ready to tango. Front Immunol. 8, 1905 (2018).
11. Perez-Nieves, B. G. et al. Dissecting phenotypic traits linked to human resilience to Alzheimer’s pathology. Brain 136, 2510–2536 (2013).
12. Huang, P. et al. REST rs7956529 genotype and rate of functional deterioration in Alzheimer’s disease. Aging Dis. 10, 94–101 (2019).
13. Selkoe, D. J. Alzheimer disease and aducanumab: adjusting our approach. Nat. Rev. Neuro. 15, 365–366 (2019).
14. Panza, F. et al. Are antibodies directed against amyloid-β (Aβ) oligomers the last call for the Aβ hypothesis of Alzheimer’s disease? Immunotherapy. 11, 3–6 (2019).
15. Louiseau, A. et al. Structural and functional features of central nervous system lymphatic vessels. Nature 523, 337–341 (2015).
16. Weisová, P. et al. Therapeutic antibody targeting microtubule-binding domain prevents neuronal internalization of extracellular tau via masking neuron surface proteoglycans. Acta Neuropathol. Commun. 7, 129 (2019).
17. Mandler, M. et al. Effects of single and combined immunotherapy approach targeting amyloid β protein and α-synuclein in a dementia with Lewy bodies-like model. Alzheimers Dement. 51552–5260, 30071–30078 (2019).
18. Heneka, M. T., O’Banion, M. K., Tervel, D. & Kummer, M. P. Neuroinflammatory processes in Alzheimer’s disease. J. Neurotraum. 117, 919–947 (2010).
19. Heneka, M. T. & O’Banion, M. K. Inflammatory processes in Alzheimer’s disease. J. Neuroimmunol. 184, 69–91 (2007).
20. Wu, J. & Li, L. Auto-antibodies in Alzheimer’s disease: potential biomarkers, pathogenic roles, and therapeutic implications. J. Biomed. Res. 30, 361–372 (2016).
21. Chitnis, T. & Weiner, H. L. CNS inflammation and neurodegeneration. J. Clin. Invest 127, 3577–3587 (2017).
22. Suzuki, K., Iwata, A. & Iwatsubo, T. The past, present, and future of disease-modifying therapies for Alzheimer’s disease. Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. 93, 757–771 (2017).
23. Herline, K., Drummond, E. & Wisniewski, T. Recent advancements toward therapeutic vaccines against Alzheimer’s disease. Expert Rev. Vaccines 17, 707–721 (2018).
24. Di Benedetto, G. et al. Beneficial effects of curtailing immune susceptibility in an Alzheimer’s disease model. J. Neuroinflamm. 16, 166 (2019).
25. Menezes, M. F. Early-onset Alzheimer disease and its variants. Continuum 25, 34–51 (2019).
26. Gao, Y. et al. Mutation profile of APP, PSEN1, and PSEN2 in Chinese familial Alzheimer’s disease. Neurobiol. Aging 77, 154–157 (2019).
27. Wingo, T, S. et al. Association of early-onset Alzheimer disease with elevated low-density lipoprotein cholesterol levels and rare genetic coding variants of APOE. JAMA Neurol. 76, 809–817 (2019).