

## ARTICLE OPEN



# The HLA-DRB1\*09:01-DQB1\*03:03 haplotype is associated with the risk for late-onset Alzheimer's disease in *APOE* $\epsilon$ 4-negative Japanese adults

Daichi Shigemizu<sup>1,2,3</sup>, Koya Fukunaga<sup>2</sup>, Akiko Yamakawa<sup>1</sup>, Mutsumi Suganuma<sup>1</sup>, Kosuke Fujita<sup>4,5</sup>, Tetsuaki Kimura<sup>1</sup>, Ken Watanabe<sup>6</sup>, Taisei Mushiroda<sup>2</sup>, Takashi Sakurai<sup>4</sup>, Shumpei Niida<sup>7</sup> and Kouichi Ozaki<sup>1,2,3</sup>

Late-onset Alzheimer's disease (LOAD) is the most common cause of dementia among those older than 65 years. The onset of LOAD is influenced by neuroinflammation. The human leukocyte antigen (HLA) system is involved in regulating inflammatory responses. Numerous HLA alleles and their haplotypes have shown varying associations with LOAD in diverse populations, yet their impact on the Japanese population remains to be elucidated. Here, we conducted a comprehensive investigation into the associations between LOAD and HLA alleles within the Japanese population. Using whole-genome sequencing (WGS) data from 303 LOAD patients and 1717 cognitively normal (CN) controls, we identified four-digit HLA class I alleles (A, B, and C) and class II alleles (DRB1, DQB1, and DPB1). We found a significant association between the HLA-DRB1\*09:01-DQB1\*03:03 haplotype and LOAD risk in *APOE*  $\epsilon$ 4-negative samples (odds ratio = 1.81, 95% confidence interval = 1.38–2.38,  $P = 2.03 \times 10^{-5}$ ). These alleles not only showed distinctive frequencies specific to East Asians but demonstrated a high degree of linkage disequilibrium in *APOE*  $\epsilon$ 4-negative samples ( $r^2 = 0.88$ ). Because HLA class II molecules interact with T-cell receptors (TCRs), we explored potential disparities in the diversities of TCR  $\alpha$  chain (TRA) and  $\beta$  chain (TRB) repertoires between *APOE*  $\epsilon$ 4-negative LOAD and CN samples. Lower diversity of TRA repertoires was associated with LOAD in *APOE*  $\epsilon$ 4-negative samples, irrespective of the HLA DRB1\*09:01-DQB1\*03:03 haplotype. Our study enhances the understanding of the etiology of LOAD in the Japanese population and provides new insights into the underlying mechanisms of its pathogenesis.

npj Aging (2024)10:3; <https://doi.org/10.1038/s41514-023-00131-3>

## INTRODUCTION

Alzheimer's disease (AD) is the most common cause of dementia among the elderly population, and it is the most frequent multifactorial neurodegenerative disease<sup>1,2</sup>. Based on the age of onset, AD is classified into two types: early-onset AD (EOAD, onset < 65 years) and late-onset AD (LOAD, onset  $\geq$  65 years). The majority of AD cases are sporadic LOAD<sup>3</sup>, which presents as a heterogeneous disorder influenced by complex interactions of both genetic and environmental risk factors. The genetic heritability is estimated to be substantial, ranging between 60% and 80%<sup>4</sup>.

Genetic associations provide valuable insights into understanding the complex etiology of this disease. Genome-wide association studies (GWAS) and WGS studies have identified susceptibility loci linked to LOAD<sup>5,6</sup>, as well as rare variants associated with LOAD<sup>7,8</sup>. LOAD-associated variants are enriched in genes like *APOE*, *ABCA7*, *BIN1*, *CR1*, *CD33*, *CLU*, *CD2AP*, *PICALM*, *SORL1*, *TREM2*, and more<sup>9</sup>. These LOAD risk genes are associated with functions ranging from immune response (*CLU*, *CR1*, *CD33*, *ABCA7*, and *TREM2*) to endocytosis (*BIN1*, *PICALM*, and *CD2AP*) and lipid processing (*APOE*, *ABCA7*, and *SORL1*)<sup>10,11</sup>. Consequently, a substantial proportion of LOAD risk genes are intertwined with the immune system. Also, our recent GWAS study has identified two distinct groups of LOAD patients, with one group characterized by immune-related genes<sup>12</sup>. Moreover, neuroinflammation has

emerged as a significant contributor to the onset and progression of LOAD<sup>10,13</sup>. Recent clinical observations have highlighted the involvement of neutrophils, essential components of the acute inflammatory response, in LOAD pathogenesis and cognitive impairment<sup>14,15</sup>.

The HLA system plays a pivotal role as a regulator of inflammatory responses, contributing to the understanding of LOAD-associated neuroinflammation<sup>16</sup>. The HLA complex is located on the short arm of chromosome 6 (6p21), which spans approximately 4000 kilobases of DNA containing more than 200 genes<sup>17</sup>. This region is divided into three major classes: class I (HLA-A, HLA-B, HLA-C), class II (HLA-DP, HLA-DQ, and HLA-DR), and class III (with most genes having unknown function)<sup>18</sup>. Numerous association studies have reported significant associations between HLA loci within the HLA class I and class II regions and LOAD. For instance, HLA-A2 alleles have been associated with LOAD risk by influencing hippocampal volume alterations<sup>19</sup>. HLA-A1 alleles and HLA-A24 alleles have been associated with delayed LOAD development in *APOE*  $\epsilon$ 4-positive adults in Italy<sup>20</sup> and *APOE*  $\epsilon$ 4-negative members of the Han Chinese population<sup>21</sup>, respectively. HLA-DRB1 alleles (specifically DRB1\*03 and DRB1\*09) and HLA-DQB1\*06 alleles have been reported to be associated with LOAD risk in German<sup>22</sup> and Iranian<sup>23</sup> populations, respectively. In addition, many HLA genes are inherited as haplotypes with varying frequencies in different populations<sup>24</sup>, such as the

<sup>1</sup>Medical Genome Center, Research Institute, National Center for Geriatrics and Gerontology, Obu, Aichi 474-8511, Japan. <sup>2</sup>RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan. <sup>3</sup>Department of Cardiovascular Medicine, Hiroshima University Graduate School of Biomedical and Health Sciences, Hiroshima 734-8551, Japan. <sup>4</sup>Department of Prevention and Care Science, Research Institute, National Center for Geriatrics and Gerontology, Obu, Aichi 474-8511, Japan. <sup>5</sup>Japan Society for the Promotion of Science, Tokyo 102-0083, Japan. <sup>6</sup>NCGG Biobank, National Center for Geriatrics and Gerontology, Obu, Aichi 474-8511, Japan. <sup>7</sup>Core Facility Administration, Research Institute, National Center for Geriatrics and Gerontology, Aichi 474-8511, Japan. ✉email: daichi@ncgg.go.jp; ozakikk@ncgg.go.jp

HLA-DRB1\*15:01-DQB1\*06:02 and HLA-DRB1\*04:02-DQB1\*03:02 haplotypes in Tunisians<sup>25</sup> and the HLA-A\*03:01-B\*07:02-DRB1\*15:01-DQA1\*01:02-DQB1\*06:02 haplotype in Europeans<sup>24</sup>. Despite these findings, the associations between HLA alleles and LOAD risk in the Japanese population have yet to be established.

Here, we conducted a comprehensive investigation into the associations between LOAD and HLA alleles within the Japanese population. Using WGS data from a large number of older Japanese people, we identified four-digit HLA class I alleles (A, B, and C) and class II alleles (DRB1, DQB1, and DPB1). Our analysis revealed an association between the HLA-DRB1\*09:01-DQB1\*03:03 haplotype and LOAD risk in *APOE*  $\epsilon$ 4-negative samples. These alleles not only showed distinctive frequencies specific to East Asians but demonstrated a high degree of linkage disequilibrium in *APOE*  $\epsilon$ 4-negative samples. Because HLA class II molecules interact with TCRs, we further explored potential disparities in the diversities of TRA and TRB repertoires between samples from *APOE*  $\epsilon$ 4-negative LOAD and CN adults. We revealed that lower diversity of TRA repertoires was associated with LOAD in *APOE*  $\epsilon$ 4-negative samples, irrespective of the HLA DRB1\*09:01-DQB1\*03:03 haplotype.

**Table 1.** Characteristics of patients whose samples were used for WGS.

Type	Phenotype	Number of samples	Age (mean $\pm$ 1 SD)	Female:male
All	LOAD	303	74.6 $\pm$ 6.3	1.53:1
	CN	1717	76.7 $\pm$ 3.0	1.14:1
<i>APOE</i> $\epsilon$ 4-positive	LOAD	108	75.2 $\pm$ 7.0	1.63:1
	CN	326	76.5 $\pm$ 3.7	1.01:1
<i>APOE</i> $\epsilon$ 4-negative	LOAD	195	74.3 $\pm$ 5.8	1.47:1
	CN	1391	76.8 $\pm$ 4.0	1.18:1

CN cognitively normal controls, LOAD late-onset Alzheimer's disease, SD standard deviation.

## RESULTS

### WGS subjects

A total of 2020 samples from Japanese donors, consisting of 303 LOAD patients and 1717 CN controls, were used in this study. All of the donors were aged at least 65 years. Ages were similar between groups, with a mean of 75 years for the LOAD group and 77 years for the CN group. The donors were further divided into whether they were positive or negative *APOE*  $\epsilon$ 4 allele. Each category had more samples from females than from males, with a large difference in the LOAD group than the CN group (Table 1).

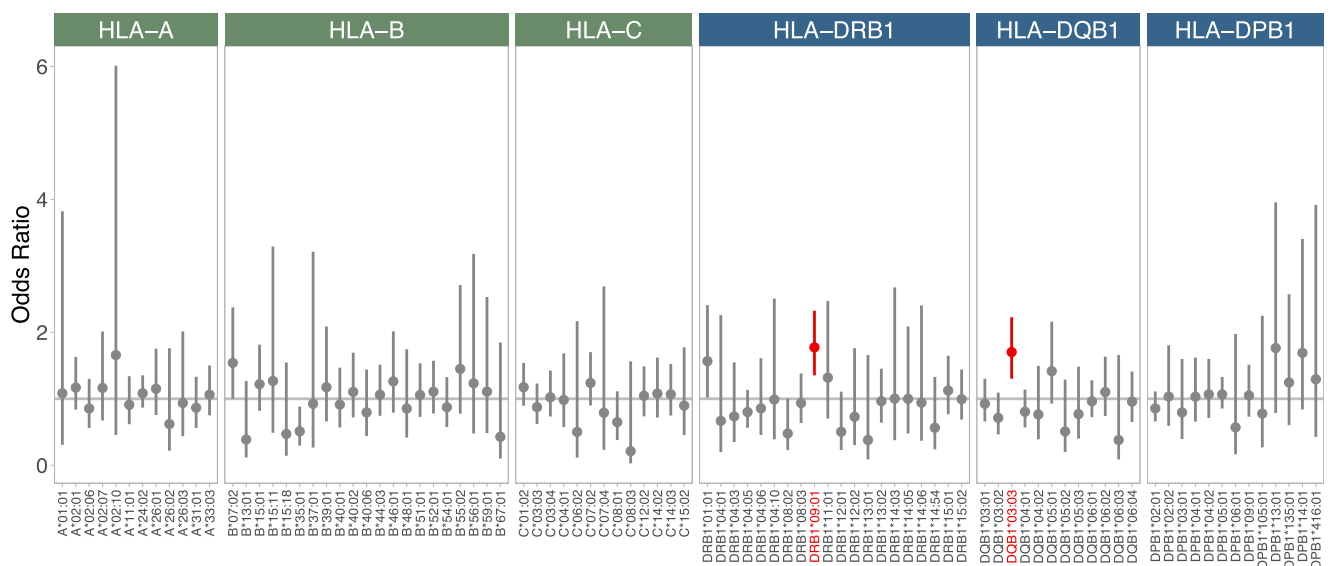
### Associations of HLA alleles

By using HISAT-genotype software<sup>26</sup>, we identified a total of 251 four-digit HLA alleles from the 2020 WGS samples. Among these alleles, 93 were classified as common alleles with allele frequency  $\geq$  0.01 (HLA-A = 13 alleles, HLA-B = 21 alleles, HLA-C = 13 alleles, HLA-DRB1 = 20 alleles, HLA-DQB1 = 12 alleles, and HLA-DPB1 = 14 alleles). Associations of these alleles with LOAD were assessed by using logistic regression, adjusting for sex and age. However, none of these alleles demonstrated a statistically significant difference in allele frequency between LOAD and CN subjects (Supplementary Figure 1).

We next examined the associations of the alleles separately for *APOE*  $\epsilon$ 4-positive and *APOE*  $\epsilon$ 4-negative subjects. While no statistically significant difference between LOAD and CN subjects were observed among *APOE*  $\epsilon$ 4-positive subjects (Supplementary Figure 2), a distinct pattern emerged for *APOE*  $\epsilon$ 4-negative subjects. We discovered that the allele frequencies of HLA-DRB1\*09:01 and HLA-DQB1\*03:03 were significantly higher in LOAD patients than in CN subjects with an adjusted *P* value  $<$  0.05 (Fig. 1 and Table 2). Neither the HLA-DRB1\*09:01 nor the HLA-DQB1\*03:03 allele showed significant deviations from Hardy-Weinberg equilibrium, as indicated by *P* values of 0.27 and 0.51, respectively.

### Validation of HLA alleles and allele frequencies of the alleles

To validate the accuracy of next-generation sequencing (NGS)-based HLA alleles, genotyping for the HLA-DRB1:09:01 allele was conducted through experimental validation at the HLA Laboratory, NPO (Kyoto, Japan). Twenty-five samples, comprising 10



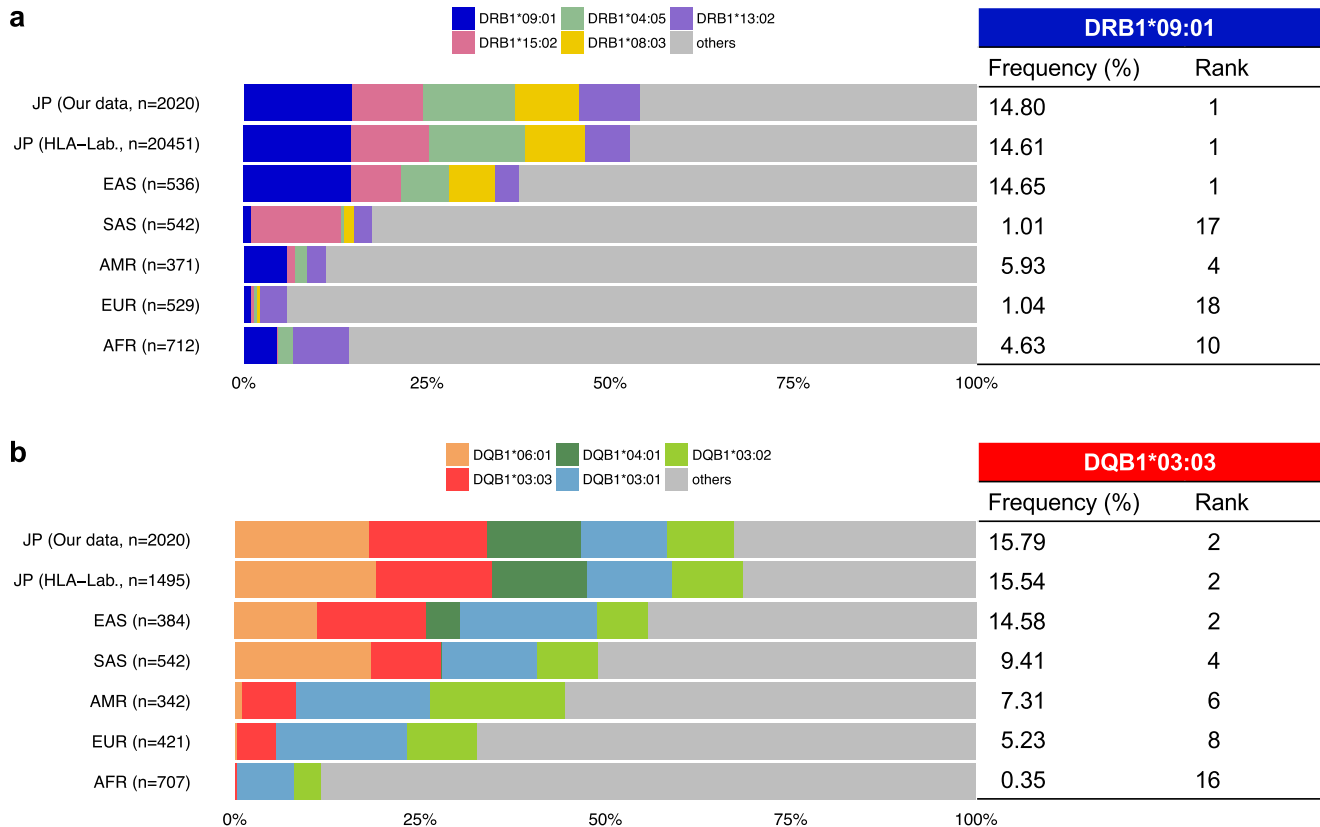
**Fig. 1** Associations of HLA class I and class II alleles with LOAD in *APOE*  $\epsilon$ 4-negative subjects. HLA class I (A, B, and C) and class II (DRB1, DQB1, and DPB1) alleles were obtained. The associations of high-resolution four-digit HLA alleles were assessed by using logistic regression with adjustment for sex and age. A false discovery rate (FDR) was calculated by using the Benjamini–Hochberg method. Red bars represent HLA alleles at an FDR  $<$  0.05. Error bars represent 95% confidence intervals.

**Table 2.** Frequencies of HLA-DRB1\*09:01 and HLA-DQB1\*03:03 alleles.

Subjects	HLA alleles (A1)	Number of samples (A1/A1A2/A2)		A1 AF		OR	95% CI	<i>P</i> *	<i>P</i> <sub>adj</sub>
		LOAD	CN	LOAD	CN				
<i>APOE</i> ε4-negative	DRB1*09:01	13/58/124	27/331/1033	0.22	0.14	1.77	1.35–2.33	$3.32 \times 10^{-5}$	$8.63 \times 10^{-4}$
	DQB1*03:03	13/60/122	30/355/1006	0.22	0.15	1.70	1.30–2.23	0.0001	0.001
<i>APOE</i> ε4-positive	DRB1*09:01	1/25/82	7/88/231	0.13	0.16	0.82	0.51–1.31	0.40	0.57
	DQB1*03:03	1/24/83	9/93/224	0.12	0.17	0.71	0.45–1.15	0.16	0.25

A2 non-A1 allele, AF allele frequency, LOAD Late-onset Alzheimer's disease, CN cognitively normal controls, OR odds ratio of LOAD vs. CN, CI confidence interval, *P*<sub>adj</sub> adjusted *P* value.

\**P* values were obtained from logistic regression with adjustment for age and sex.



**Fig. 2** Allele frequencies of HLA-DRB1\*09:01 and HLA-DQB1\*03:03 in several populations. Allele frequencies of HLA-DRB1\*09:01 (a) and HLA-DQB1\*03:03 (b) in six populations were obtained from three datasets: our WGS data, the HLA Laboratory, and the 1000 Genome Project. The five top-ranked HLA-DRB1 (a) and HLA-DQB1 (b) alleles in our WGS data are represented, and the HLA-DRB1\*09:01 (a) and HLA-DQB1\*03:03 (b) are shown in blue and red, respectively. JP Japanese; EAS East Asian, SAS South Asian, AMR admixed American, EUR European, AFR African, n number of samples.

homozygotes and 10 heterozygotes possessing the HLA allele of interest, along with five homozygotes lacking the HLA allele, were randomly selected from *APOE* ε4-negative samples. Of the 50 alleles assessed, 49 were consistent with our NGS-based HLA alleles, resulting in a high concordance rate of 0.98 (Supplementary Table 1).

The allele frequencies of HLA-DRB1\*09:01 and HLA-DQB1\*03:03 were found to be 14.80% and 15.79%, respectively, within our Japanese WGS dataset. These frequencies closely matched those obtained from a substantial collection of healthy Japanese samples provided by the HLA Laboratory, NPO (Kyoto, Japan) (HLA-DRB1\*09:01 = 14.61%, HLA-DQB1\*03:03 = 15.54%, Fig. 2). We also compared our results with allele frequencies from various populations in data provided by the 1000 Genomes Project. Both HLA alleles, DRB1\*09:01 and DQB1\*03:03 had a high occurrence

rate, particularly within the East Asian population, including Japan. DRB1\*09:01 ranked first in this population, and DQB1\*03:03 ranked second (Fig. 2).

#### HLA-DRB1\*09:01-DQB1\*03:03 haplotype

Since the allele frequencies of HLA-DRB1\*09:01 in LOAD and CN subjects showed similar frequencies of those of HLA-DQB1\*03:03 in LOAD and CN subjects (Table 2), we examined the linkage disequilibrium (LD) of these alleles. Our finding revealed a high LD within *APOE* ε4-negative subjects, with  $r^2$  values of 0.88. Subsequently, we examined whether the haplotype formed by these alleles was associated with the risk for LOAD. Our analysis yielded significant results, with the haplotype's allele frequency being considerably higher in LOAD patients than in CN subjects (odds ratio = 1.81, 95% confidence interval = 1.38 to 2.38,

**Table 3.** Frequencies of the HLA-DRB1\*09:01-DQB1\*03:03 haplotype.

Subjects	Number of samples (A1/A1A2/A2) A1 AF		OR	95% CI	P*		
	LOAD	CN				LOAD	CN
<i>APOE</i> $\epsilon$ 4-negative	13/57/125	25/326/1040	0.21	0.14	1.81	1.38–2.38	$2.03 \times 10^{-5}$
<i>APOE</i> $\epsilon$ 4-positive	1/24/83	7/87/232	0.12	0.16	0.79	0.49–1.28	0.34

A1 HLA-DRB1\*09:01-DQB1\*03:03 haplotype, A2 non-A1 haplotype, AF allele frequency, LOAD Late-onset Alzheimer's disease, CN cognitively normal controls, OR odds ratio of LOAD vs. CN, CI confidence interval.

\*P value was obtained from logistic regression with adjustment for age and sex.

**Table 4.** Clonal diversity of TCR repertoires between phenotypes in *APOE*  $\epsilon$ 4-negative subjects.

TCR type	Adjustment	OR	95% CI	P*
TRA	Age + Sex	0.983	0.966–0.998	0.034
	Age + Sex + haplotype <sup>a</sup>	0.982	0.966–0.998	0.033
TRB	Age + Sex	0.984	0.965–1.002	0.097
	Age + Sex + haplotype <sup>a</sup>	0.995	0.966–1.003	0.11

TCR T-cell receptor, TRA T-cell receptor alpha, TRB T-cell receptor beta, OR odds ratio of LOAD vs. CN, CI confidence interval.

\*P values were obtained from logistic regression.

<sup>a</sup>Haplotype: HLA-DRB1\*09:01-DQB1\*03:03.

Table 3). Furthermore, we examined the associations between the number of haplotypes and Mini-Mental State Examination (MMSE) scores, separately for *APOE*  $\epsilon$ 4-positive and *APOE*  $\epsilon$ 4-negative LOAD patients. The associations were evaluated by using linear regression analysis, with adjustments for sex and age. However, we found no statistically significant differences between them (*APOE*  $\epsilon$ 4-positive LOADs,  $P = 0.96$ ; *APOE*  $\epsilon$ 4-negative LOADs,  $P = 0.12$ ).

Kawabata et al. previously reported that people homozygous for the DRB1\*09:01-DQB1\*03:03 haplotype were more frequent among Japanese with type 1 diabetes than in control subjects<sup>27</sup>. We therefore checked whether this specific haplotype was also associated with diabetes in *APOE*  $\epsilon$ 4-negative Japanese people. We selected a subset of 1586 samples from *APOE*  $\epsilon$ 4-negative subjects, including 53 patients with diabetes and 166 without diabetes (controls) according to defined diabetes criteria (see "Methods"). Within the diabetes patient group, there were three haplotype homozygotes, 21 haplotype heterozygotes, and 29 non-haplotype homozygotes, although the diabetes patients aged  $\geq 65$  years were more likely to have type 2 diabetes. The control group consisted of 10 haplotype homozygotes, 40 haplotype heterozygotes, and 116 non-haplotype homozygotes. We conducted an association analysis between the haplotype and diabetes/control samples by using Fisher's exact test. However, there was no statistically significant association in recessive model ( $P > 0.99$ , Supplementary Table 2). We concluded that the HLA-DRB1\*09:01-DQB1\*03:03 haplotype was not associated with diabetes within our *APOE*  $\epsilon$ 4-negative Japanese population.

### Clonal diversity of TCR repertoires

HLA class II molecules, including the HLA-DRB1 and HLA-DQB1 variants present in our haplotype, play a crucial role in interacting with TCRs. Given this interaction, we explored the potential impact of our identified haplotype on the diversities of TCR repertoires within *APOE*  $\epsilon$ 4-negative Japanese adults. TRUST4 software<sup>28</sup> was used to estimate the diversities of TRA and TRB repertoires from bulk RNA-sequencing (RNA-seq) within each sample (see "Methods"). Our dataset comprised 114 RNA-seq samples derived from *APOE*  $\epsilon$ 4-negative subjects (55 LOAD and 59 CN samples) obtained

from the National Center for Geriatrics and Gerontology (NCGG) Biobank database. A logistic regression model was used to identify statistically significant increases or decreases in the clonal diversity of TCR repertoires between phenotypes with adjustments for age and sex and for age, sex, and our specific haplotype (i.e., DRB1\*09:01-DQB1\*03:03). Despite the lack of statistically significant differences in the TRB repertoires under either adjustment, the diversity of the TRA repertoires did demonstrate statistically significant decreases under both adjustments in *APOE*  $\epsilon$ 4-negative samples (Table 4). These findings provided insights into the associations between the diversities of TRA repertoires and the risk for LOAD within the *APOE*  $\epsilon$ 4-negative Japanese population, regardless of the presence of the HLA DRB1\*09:01-DQB1\*03:03 haplotype.

### DISCUSSION

Through our investigation involving WGS data from 2020 Japanese people, a specific HLA class II haplotype, HLA-DRB1\*09:01-DQB1\*03:03, emerged as being associated with the underlying mechanisms of LOAD. Although both HLA-DRB1 and HLA-DQB1 alleles are commonly observed in East Asian populations, their allele frequencies vary across different ethnic populations. Within the HLA-DQB1 locus, the HLA-DQB1\*03:01 allele is frequently observed across all populations, whereas the HLA-DQB1\*06:01 allele is particularly common among Asian populations. Our identified HLA-DQB1\*03:03 allele had the second highest allele frequency after the HLA-DQB1\*06:01 allele among Japanese adults. In the HLA-DRB1 locus, the HLA-DRB1\*09:01 allele was the most frequent within East Asian populations, accounting for approximately 15%. This stands in contrast to South Asian, American, and European populations, where the HLA-DRB1\*07:01 allele is more prominent (SAS = 19%, AMR = 10%, EUR = 14%). In Africans, HLA-DRB1\*15:03 was the most frequent, accounting for approximately 12%. The HLA-DRB1\*15:03 allele, dominant in Africans, is not present in East Asian, South Asian, or European populations. Many LOAD-associated HLA alleles have been reported in similar studies<sup>20,21,24,25</sup>, but our findings were not included in their results due to their East Asian specificity. This result indicates the significance of population-specific variations in HLA allele frequencies and highlights the importance of region-specific investigations in understanding complex diseases like LOAD.

Various methods for HLA genotyping using NGS data have been developed<sup>26,29,30</sup> and evaluated for their concordance rates, particularly at 2-digit and 4-digit allele resolutions<sup>31</sup>. In this study, we employed a highly accurate NGS-based HLA genotyping method called HISAT-genotype. Although the method has demonstrated an extremely high concordance rate ( $\geq 0.97$ ) at the 2-digit allele level (e.g., HLA-A\*01), the concordance rate at the 4-digit allele resolution has been slightly lower, especially for the HLA-DRB1 gene (with a concordance rate of 0.87)<sup>26</sup>. Because our findings included an HLA-DRB1 locus, we experimentally validated our HLA-DRB1 alleles. The concordance rate of these alleles was significantly higher than we expected, reaching 0.98. This

outcome provides robust support for the reliability and validity of NGS-based HLA genotyping, suggesting its suitability for potential application in clinical use.

Our study has revealed that the DRB1\*09:01-DQB1\*03:03 haplotype is associated with the risk of LOAD in *APOE*  $\epsilon$ 4-negative Japanese adults. This haplotype has previously been linked to certain diseases in the Japanese population, such as type 1 diabetes<sup>27</sup> and myeloperoxidase-antineutrophil cytoplasmic antibody-positive vasculitis (MPO-AAV)<sup>32</sup>. Although *APOE*  $\epsilon$ 4-negative Japanese diabetes patients aged  $\geq 65$  years were more likely to have type 2 diabetes, our findings did not show a statistically significant association between the haplotype and the diabetes in *APOE*  $\epsilon$ 4-negative Japanese people. However, further HLA genotyping and association studies in a larger number of samples may reveal a relationship between the HLA haplotype and diabetes risk. MPO-AAV is a life-threatening autoimmune disease which causes severe inflammation and destruction of small blood vessels, mainly in the kidneys of people older 50 years<sup>33</sup>. Given that neuroinflammation significantly contributes to the onset and progression of LOAD<sup>10,13</sup>, this result suggests that the presence of the DRB1\*09:01-DQB1\*03:03 haplotype might confer an increased risk of LOAD, even in adults without the *APOE*  $\epsilon$ 4 allele.

Indeed, our investigation into the interaction between the DRB1\*09:01-DQB1\*03:03 haplotype and the diversity of TCR repertoires indicated that a low diversity of TRA repertoires is associated with LOAD onset in *APOE*  $\epsilon$ 4-negative Japanese adults, providing valuable insights into the potential mechanisms underlying the association between this haplotype and LOAD risk in *APOE*  $\epsilon$ 4-negative Japanese people. Xu et al. previously reported that the diversity of TCR repertoires was significantly lower in samples from AD patients than in samples from CN individuals, as analyzed with single-cell RNA-seq<sup>34</sup>. Our previous study also reported that the diversity of B-cell receptor repertoires (IGH and IGK) and TRA were significantly smaller in people with LOAD progression, using a large number of RNA-seq datasets ( $> 850$  Japanese samples)<sup>35</sup>. These results suggest that the DRB1\*09:01-DQB1\*03:03 haplotype and the diversity of TCR repertoires could potentially be associated with the onset of LOAD through different mechanisms. However, given that HLA Class II molecules interact with CD4+ helper T-cells, we may need to investigate the association between this specific haplotype and the diversity of specific CD4+ helper TCR repertoires in *APOE*  $\epsilon$ 4-negative Japanese adults in the near future.

To our knowledge, this study represents the largest NGS-based HLA genotyping analysis in Japanese LOAD cases. However, further investigations using larger sample sizes are likely to reveal additional pathogenic HLA alleles associated with LOAD. Our findings contribute to enhancing the understanding of LOAD and provide insight into its pathogenic mechanisms for future investigations. While our current dataset may not yet be sufficient to comprehensively identify all HLA alleles associated with LOAD in the Japanese population, undertaking an association study between LOAD and CN through a large number of WGS samples will undoubtedly lead to the discovery of new HLA alleles associated with LOAD.

## METHODS

### Ethics statements

This study received approval from the ethics committee of the National Center for Geriatrics and Gerontology. The design and execution of the current study, which involved human subjects, were clearly described in a research protocol. Participation in the NCGG Biobank was voluntary, and all participants provided written informed consent before enrolling. Surrogate consent was obtained for patients with advanced cognitive decline.

## Subjects

All of the blood samples used in this study and the associated clinical data were obtained from the NCGG Biobank. Of the 2020 samples, 303 were from LOAD patients and the remaining 1717 were from CN subjects. Donors with LOAD were diagnosed with AD according to the criteria established by the National Institute on Aging—Alzheimer's Association workgroups<sup>36,37</sup>. CN subjects had subjective cognitive complaints but normal cognition on a neuropsychological assessment with a comprehensive neuropsychological test and a MMSE score  $> 23$ . Diagnosis of all AD in all donors whose samples were used was based on medical history, physical examination and diagnostic tests, neurological examination, neuropsychological tests, and brain imaging with magnetic resonance imaging or computerized tomography. Diagnoses were made by experts, including neurologists, psychiatrists, geriatricians, and neurosurgeons, all of whom were experts in dementia and familiar with its diagnostic criteria. All of the samples were from adults 65 years of age or older.

## WGS data analysis

DNA concentration was measured by using a PicoGreen DNA assay, and fragmentation of DNA was assessed with agarose gel electrophoresis. High-quality DNA was used for DNA libraries. A WGS library was constructed by using the TruSeq DNA PCR-Free Library Preparation Kit (Illumina, Inc., San Diego, CA, USA) in accordance with the manufacturer's instructions. WGS was conducted at Macrogen Japan Corporation, Takara Bio Inc., and GENEWIZ Inc. The DNA was sequenced by using either the Illumina HiSeq X Ten or NovaSeq 6000 platform, generating paired-end reads of 151 base pairs in accordance with the manufacturer's instructions. All data used in this study were obtained from the NCGG Biobank database<sup>8</sup>, where they were initially registered.

## HLA genotyping from WGS data

NGS-based HLA genotyping was performed by using HISAT-genotype software<sup>26</sup>, which is accessible through a public GitHub repository (1.3.2 release; <https://daehwankimlab.github.io/hisat-genotype/>). Individual HLA genotyping was conducted by using two programs: "hisatgenotype" and "hisatgenotype\_toolkit". These programs facilitated the determination of HLA class I alleles (A, B, and C) and class II alleles (DRB1, DQB1, and DPB1). For case-control association studies, the high-resolution four-digit HLA alleles were employed. The associations were evaluated by using logistic regression analysis, with adjustments for sex and age. Adjusted *P* values were calculated with the Benjamini–Hochberg method. An adjusted *P* value of 0.05 or less was considered statistically significant. Logistic regression analysis (–logistic) and Hardy–Weinberg equilibrium test (–hardy) for HLA loci were performed by using PLINK software (version 1.90b)<sup>38</sup>. Allele frequencies of HLA loci in various populations were available publicly at [https://hla.or.jp/med/frequency\\_search/ja/allele/](https://hla.or.jp/med/frequency_search/ja/allele/) (the HLA Laboratory, NPO) and at [http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data\\_collections/HLA\\_types/](http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/HLA_types/) (the 1000 Genomes Project).

## HLA haplotype

Beagle software (version 3.0.4)<sup>39</sup> was used to phase HLA alleles into individual HLA haplotypes. This process was facilitated by employing two programs, "linkage2beagle.jar" and "beagle.jar". The software package can be downloaded from [http://faculty.washington.edu/browning/beagle/recent.versions/beagle\\_3.0.4\\_05May09.zip](http://faculty.washington.edu/browning/beagle/recent.versions/beagle_3.0.4_05May09.zip). To assess the degree of linkage disequilibrium between two HLA loci,  $r^2$  values were calculated by using PLINK software and the (–ld) parameter.

## Incident diabetes

Incident diabetes was identified through the following criteria: self-report of diabetes, usage of antihyperglycemic medicines, or hemoglobin A1c (HbA1c) levels  $\geq 6.5\%$ <sup>40</sup>. Conversely, subjects whose records lacked self-report of diabetes and antihyperglycemic medicine use while displaying HbA1c  $< 6.5\%$ , were classified as not having diabetes (i.e., control samples). The examination of haplotype associations between diabetes and control samples were carried out with Fisher's exact test, facilitated by the PLINK software with the ( $-$ model fisher) parameter<sup>38</sup>.

## Detection of T-cell receptor repertoires

All RNA-seq data were downloaded from the NCGG Biobank database<sup>15</sup>. TCR repertoires were identified from RNA-seq data using TRUST4 software (v1.0.5). This included the inference of CDR3 clonotypes, encompassing both the TRA and TRB. To estimate the clonal diversity of TCRs, an inverse Simpson index was employed, and the calculation was conducted by using VDJtools<sup>41</sup> (v1.2.1). Assessments regarding the associations of the clonal diversity within TCR repertoires were conducted through logistic regression analysis between phenotypes, which were implemented with the R programming language (R Development Core Team, <http://www.r-project.org/>).

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## DATA AVAILABILITY

All datasets used or analyzed are available from the corresponding author on reasonable request.

## CODE AVAILABILITY

We used the open-source program languages R (version 3.4.1) and Ruby (version 2.4.0) to analyze data and create plots. Code is available upon request from the corresponding authors.

Received: 27 September 2023; Accepted: 22 November 2023;

Published online: 02 January 2024

## REFERENCES

- Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).
- Prince, M. et al. The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimers Dement* **9**, 63–75.e62 (2013).
- Rabinovici, G. D. Late-onset Alzheimer disease. *Continuum (Minneapolis Minn)* **25**, 14–33 (2019).
- Gatz, M. et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* **63**, 168–174 (2006).
- Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
- Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
- Prokopenko, D. et al. Whole-genome sequencing reveals new Alzheimer's disease-associated rare variants in loci related to synaptic function and neuronal development. *Alzheimers Dement* **17**, 1509–1527 (2021).
- Shigemizu, D. et al. Whole-genome sequencing reveals novel ethnicity-specific rare variants associated with Alzheimer's disease. *Mol. Psychiatry* **27**, 2554–2562 (2022).
- Li, Y. et al. Genomics of Alzheimer's disease implicates the innate and adaptive immune systems. *Cell Mol. Life Sci.* **78**, 7397–7426 (2021).
- Villegas-Llerena, C., Phillips, A., Garcia-Reitboeck, P., Hardy, J. & Pocock, J. M. Microglial genes regulating neuroinflammation in the progression of Alzheimer's disease. *Curr. Opin. Neurobiol.* **36**, 74–81 (2016).
- Chan, G. et al. CD33 modulates TREM2: convergence of Alzheimer loci. *Nat. Neurosci.* **18**, 1556–1558 (2015).
- Shigemizu, D. et al. Classification and deep-learning-based prediction of Alzheimer disease subtypes by using genomic data. *Transl. Psychiatry* **13**, 232 (2023).
- Ransohoff, R. M. How neuroinflammation contributes to neurodegeneration. *Science* **353**, 777–783 (2016).
- Zenaro, E. et al. Neutrophils promote Alzheimer's disease-like pathology and cognitive decline via LFA-1 integrin. *Nat. Med.* **21**, 880–886 (2015).
- Shigemizu, D. et al. Identification of potential blood biomarkers for early diagnosis of Alzheimer's disease through RNA sequencing analysis. *Alzheimers Res. Ther.* **12**, 87 (2020).
- Heppner, F. L., Ransohoff, R. M. & Becher, B. Immune attack: the role of inflammation in Alzheimer disease. *Nat. Rev. Neurosci.* **16**, 358–372 (2015).
- Horton, R. et al. Gene map of the extended human MHC. *Nat. Rev. Genet.* **5**, 889–899 (2004).
- Choo, S. Y. The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei. Med. J.* **48**, 11–23 (2007).
- Wang, Z. X. et al. HLA-A2 alleles mediate Alzheimer's disease by altering hippocampal volume. *Mol. Neurobiol.* **54**, 2469–2476 (2017).
- Guerini, F. R. et al. HLA-A\*01 is associated with late onset of Alzheimer's disease in Italian patients. *Int. J. Immunopathol. Pharmacol.* **22**, 991–999 (2009).
- Ma, S. L. et al. Association between HLA-A alleles and Alzheimer's disease in a southern Chinese community. *Dement. Geriatr. Cogn. Disord.* **26**, 391–397 (2008).
- Neill, D. et al. Risk for Alzheimer's disease in older late-onset cases is associated with HLA-DRB1\*03. *Neurosci. Lett.* **275**, 137–140 (1999).
- Rezaei Rad, F., Ghahvechi Akbari, M., Zamani, M., Bayat, S. & Zamani, M. Pharmacogenetic and association studies on the influence of HLA alleles and rivastigmine on the Iranian patients with late-onset Alzheimer's disease. *Mol. Neurobiol.* **58**, 2792–2802 (2021).
- 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).
- Mansouri, L. et al. Association of HLA-DR/DQ polymorphism with Alzheimer's disease. *Am. J. Med. Sci.* **349**, 334–337 (2015).
- Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
- Kawabata, Y. et al. Asian-specific HLA haplotypes reveal heterogeneity of the contribution of HLA-DR and -DQ haplotypes to susceptibility to type 1 diabetes. *Diabetes* **51**, 545–551 (2002).
- Song, L. et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods* **18**, 627–630 (2021).
- Nariai, N. et al. HLA-VBSeq: accurate HLA typing at full resolution from whole-genome sequencing data. *BMC Genomics* **16**, 57 (2015).
- Ka, S. et al. HLAScan: genotyping of the HLA region using next-generation sequencing data. *BMC Bioinformatics* **18**, 258 (2017).
- Liu, P. et al. Benchmarking the human leukocyte antigen typing performance of three assays and seven next-generation sequencing-based algorithms. *Front. Immunol.* **12**, 652258 (2021).
- Kawasaki, A. et al. Association of HLA-class II alleles with risk of relapse in myeloperoxidase-antineutrophil cytoplasmic antibody positive vasculitis in the Japanese population. *Front. Immunol.* **14**, 1119064 (2023).
- Odobasic, D. & Holdsworth, S. R. Emerging cellular therapies for anti-myeloperoxidase vasculitis and other autoimmune diseases. *Front. Immunol.* **12**, 642127 (2021).
- Xu, H. & Jia, J. Single-cell RNA sequencing of peripheral blood reveals immune cell signatures in Alzheimer's disease. *Front. Immunol.* **12**, 645666 (2021).
- Shigemizu, D., Akiyama, S., Mitsumori, R., Niida, S. & Ozaki, K. Identification of potential blood biomarkers for early diagnosis of Alzheimer's disease through immune landscape analysis. *NPJ Aging* **8**, 15 (2022).
- McKhann, G. M. et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 263–269 (2011).
- Albert, M. S. et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 270–279 (2011).
- Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
- Lipska, K. J. et al. Elevated HbA1c and fasting plasma glucose in predicting diabetes incidence among older adults: are two better than one? *Diabetes Care* **36**, 3923–3929 (2013).

41. Shugay, M. et al. VDJtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).

## ACKNOWLEDGEMENTS

We thank the NCGG Biobank for providing the study materials, clinical information, and technical support. This study was supported by grants from AMED (Grant Numbers JP18kk0205009 and JP18kk0205012 to S.N., Grant Numbers JP21dk0207045 and JP22dk0207060 to S.N. and K.O., Grant Numbers JP21de0107002 and JP21dk0207052 to S.N. and D.S., and Grant Number JP21km0405501 to K.O.); Research Funding for Longevity Sciences from the NCGG (21–23 to K.O., and 21–24 and 23–7 to D.S.); JSPS KAKENHI (Grant Number JP21H02470 to D.S.); and a grant from the Japanese Ministry of Health, Labour, and Welfare for Research on Dementia (to K.O.). We also thank ELSS editors for English corrections.

## AUTHOR CONTRIBUTIONS

D.S. designed the study and the main conceptual ideas and performed the analyses; A.Y., K. Fukunaga, and T.M. provided technical assistance; M.S. and T.K. planned the experiments for validation of HLA alleles; K. Fujita, K.W. and T.S. collected the data; D.S. wrote the manuscript; and D.S., S.N., and K.O. organized this work. All authors contributed to and approved the final manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41514-023-00131-3>.

**Correspondence** and requests for materials should be addressed to Daichi Shigemizu or Kouichi Ozaki.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024