Title: Stereotypic persistent B cell receptor clonotypes in Alzheimer's Disease

Hyunji Yang, Namphil Kim, Yonghee Lee, Duck Kyun Yoo, Jinny Choi, Ki Woong Kim, Jong Bin Bae, Ji Won Han, Sunghoon Kwon, Junho Chung

Corresponding author

S. Kwon (skwon@snu.ac.kr) or J. Chung (jjhchung@snu.ac.kr)

Abstract

We constructed *in silico* B cell receptor (BCR) repertoires using peripheral blood (PB) samples collected from 44 Alzheimer's Disease (AD) patients at baseline and 37 patients at follow-up. For the control group (CG), we used BCR repertoire data from the chronologically collected PB samples of 55 healthy volunteers vaccinated with the SARS-CoV2 mRNA vaccine. AD patients shared 3,983 stereotypic BCR clonotypes not found in CG, and their degree of overlap between patient pairs were significantly higher than that of CG pairs, even with the SAS-CoV2 spike protein triggering a concerted BCR response. Many stereotypic AD patient-specific BCR clonotypes co-existed in more than four patients and persisted throughout the two sampling points. One of these BCR clonotypes encoded an antibody reactive to the Aβ42 peptide. Our findings strongly suggest that AD patients are exposed to common (auto)antigens associated with disease pathology, and that their BCR repertoire signatures have high diagnostic potential.

Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that causes dementia, cognitive decline, memory loss, and impaired daily functioning. Thus far, the presence of amyloid beta peptide (Aβ) plaques and neurofibrillary tangles (NFTs) composed of phosphorylated tau proteins within the central nervous system (CNS) have been discussed as the major contributors to neuroinflammation and the pathological hallmarks of AD[1, 2]. However, recent studies have revealed that the

pathogenesis of AD is not solely restricted to these changes in the brain[3]. An increasing body of evidence points to the peripheral immune system as another major player intimately linked to AD pathology[4-7]. Several studies have demonstrated that the peripheral homeostasis of T lymphocytes is disrupted in AD patients[8-10], but little is known about the role of peripheral B lymphocytes. A single-cell RNA sequencing analysis found that the progression of AD is associated with a reduction in the number of B cells in the peripheral blood, as well as differentially expressed genes in B cells[11]. B cells play a fundamental role in the adaptive immune system particularly in humoral immunity, by recognizing antigens via membrane-expressed B cell receptors (BCR) and producing antigen-specific immunoglobulin (Ig) in response[12]. Accordingly, recent studies have investigated the differences in immune cell type compositions and the clonal diversity of BCR repertoires among AD patients through single-cell RNA sequencing data or single-cell BCR libraries of peripheral blood mononuclear cells (PBMCs)[5,13]. The results showed that the proportion of B cells and the clonal diversity of both the heavy and Kappa chains of BCR repertoires all significantly decreased with the progression of AD. Previous reports including ours have characterized the BCR repertoires of AD patients, such as isotype distribution, variable (V) and joining (J) gene usage, somatic hypermutation (SHM) patterns, and the diversity of the complementarity-determining region 3 (CDR3)[5,14]. Despite their success in highlighting the importance of BCRs in AD neuropathology, the findings were limited by the relatively small sample sizes of 3 and 10, and the lack of chronological sampling. Furthermore, they have not characterized the stereotypic BCRs at an individual sequence level, or identified their specific antigens.

In this study, we identified BCR clonotypes that are persistently present and shared among AD patients, but not in healthy populations. In order to analyze their characteristics, we collected peripheral blood (PB) from 44 AD patients at baseline and chronological PB samples from 37 patients with an interval of 14 - 53 weeks, and re-constituted their BCRs *in silico*. For the control group (CG), we used BCR repertoires from 55 healthy volunteers vaccinated with the SARS-CoV2 mRNA vaccine at baseline, whose PB samples were collected six times over 42 weeks (285 PB samples obtained chronologically)[15]. Results showed that the AD patients had a significantly higher abundance of stereotypic patient-specific BCR clonotypes compared to the vaccine-specific BCR clonotypes in the CG. In addition, one of the

identified stereotypic AD patient-specific BCR clonotypes was found to encode antibodies reactive to the Aβ42 peptide, a well-known and key antigen associated with AD progression[16].

Main text

Selection of stereotypic AD patient-specific non-naïve BCR clonotypes

For a period of 75 weeks, peripheral blood (PB) samples were collected from 44 Alzheimer patients with a average age of 75 years (standard deviation; 7.04) and sex ratio of 34% and 66% (male vs female). PB samples were collected twice with an interval of 14 to 53 weeks, with the exception of seven patients who withdrew from the study between time point (TP) 1 and 2 (Patient ID 3, 5, 8, 19, 28, 30, and 37) (Fig. 1a). The duration of the first PB sampling period was 61 weeks, and the second PB sampling 34 weeks, respectively. From these 81 PB samples, the cDNA was prepared and the gene fragment that encodes both the variable domain (VH) and the N-terminal region of the BCR heavy chain constant domain 1 (CH1) was subjected to amplification using a specific primer set (Extended Table 2). The amplified product was then analyzed by next-generation sequencing (NGS) with a median read of 245,033 (standard deviation of 184,462, Extended Table 3). For BCR repertoire analysis, we first removed naïve BCR sequences, which we define as those with the IgM or IgD isotype containing less than 1 SHM. We then grouped the remaining sequences into clonotypes, which we define as a collection of sequences encoded by identical IGHV, IGHJ genes and CDR3 sequence at the amino acid level[17]. Among these BCR clonotypes, AD patient-specific clonotypes were selected by excluding the clonotypes also found in our prior studies among non-AD individuals. This CG was constituted with 55 healthy subjects (285 PB samples obtained chronologically) injected three times with the COVID-19 vaccine[15] (Fig. 1b, Extended Table 1). The procedures described above were replicated in the CG group to collect the control non-naïve BCR clonotypes. Before further analysis, we excluded all stereotypic clonotypes that did not show unique BCR sequences at the nucleotide level, in order to account for potential aerosol molecular contamination and index switching during NGS library preparation and sequencing[15]. At the end of this

selection processes, we were able to identify 3,983 BCR clonotypes specific to the AD patient group, and 83,821 BCR clonotypes specific to the CG, respectively (Fig. 1c).

Stereotypic AD patient-specific non-naïve clonotypes showed a significantly higher degree of overlap in patient pairs compared to CG pairs.

We then analyzed the distribution of stereotypic AD patient-specific BCR clonotypes in 1,892 ((44 X 44)-44) and 1,332 ((37 X 37)-37) patient pairs in the first and second PB samples, respectively. We calculated the degree of overlap for stereotypic BCR clonotypes in each pair using the cosine similarity method, with one-hot encoded vectors representing the presence or absence of each stereotypic clonotype in a patient's BCR repertoire. We also obtained the degree of overlap for the CG pairs over the six PB sampling TPs. As expected, there was a dramatic increase in the average cosine similarity compared to the pre-vaccinated state (TP1) at TP3, TP4, and the highest increase at TP6. Each TP corresponds to immediately after the second- (TP3), 1 month after the second- (TP4), and immediately after the third-injection (TP6) of the COVID-19 vaccine. This increase in average cosine similarity be attributed to the concerted BCR response toward the COVID-19 immunogen (Fig 2a-c, $p = 4.92 \times 10^{-6}$, $p = 2.52 \times 10^{-30}$, $p = 2.26 \times 10^{-112}$). The AD patient pairs also showed significantly higher average cosine similarity at both sampling points compared to the CG's baseline (TP1) ($p = 2.06 \times 10^{-21}$, $p = 1.22 \times 10^{-169}$). The cosine similarity value of the AD patient samples was higher at TP2, possibly due to the shorter PB sampling duration (61 weeks vs 34 weeks), which would provide less diverse seasonal antigens to AD patients. Surprisingly, the average cosine similarity of the AD patients' TP2 BCR clonotypes was significantly higher than even that of the CG's BCR clonotypes at TP6 ($p = 1.86 \times 10^{-6}$). This observation is interesting, especially under the consideration that the PB sampling duration was much longer in the AD patients compared to the CG (34 weeks vs 6 weeks). To account for the fact that we are not aware of the COVID-19 infection history or vaccination status of the AD patients, we performed the same analysis on all stereotypic clonotypes identified in the two groups without excluding the stereotypic BCR clonotypes of the other. The resulting cosine similarity values of the AD patients' BCR clonotypes at TP1 and TP2

were both significantly higher than those of the CG's BCR clonotypes at all TPs except for TP6, which did not show statistically significant difference (Extended Figure 1a-c). These results clearly demonstrate that AD patients share stereotypic BCR clonotypes at an equivalent or higher degree compatible to vaccinees who received their third COVID-19 injection within six weeks prior to PB sampling, irrespective of the AD patients' longer PB sampling period.

Our analysis showed that the degree of overlap in the stereotypic AD patient-specific BCR clonotypes significantly differs among patient pairs. This suggests that AD patients can be placed in subgroups categorized by their BCR clonotypes. To assign these patient subgroups, the stereotypic AD patient-specific BCR clonotypes at all TPs in each patient were combined, and the cosine similarity of patient pairs were calculated. Twenty patient pairs with the highest cosine value were selected and displayed on a Chord plot (Fig. 2d), in which the patients were divided into six subgroups. Among these pairs, one patient pair (patient ID 26 and 28) had an outlying cosine similarity value of 0.961, and the number of BCR clonotypes was also noticeably higher than the other pairs (Fig. 2e). Indeed, between the 1,225 and 1,228 stereotypic BCR clonotypes in patient ID 26 and patient ID 28, respectively, 1,179 clonotypes were shared. The presence of these AD patient subgroups showing a high degree of overlap in their BCR clonotypes strongly suggests the presence of common and possibly auto- antigens within the subgroup.

A highly shared and persistent AD patient-specific BCR clonotype encodes an antibody reactive to the Aβ42 peptide

To identify the antigens to which these stereotypic AD patient-specific BCR clonotypes are reactive, we selected twenty clonotypes that were most highly shared between four or more patients (Fig. 3a). We also analyzed the persistence of these clonotypes over the two TPs. One clonotype encoded by the IGHV3-7/IGHJ1 genes with the CDR3 sequence VRLAEYFQN was present at both TPs in three patients. When we allowed one amino acid mismatch in the CDR3 sequences, we found an additional patient (patient ID 34) with a similar clonotype (CDR3 sequence: VRLAEYFQH) (Fig. 3a). In 285 PB samples of the CG, we did not find any BCR

clonotypes encoded by the IGHV3-7/IGHJ1 genes and with a CDR3 sequence homologous to VRLAEYFQN, even with amino acid mismatch allowance (Fig. 3b). We then sequenced all the BCR clonotypes encoded by IGHV3-7/IGHJ1 with CDR3 sequences homologous to VRLAEYFQN, given one amino acid mismatch allowance (Fig. 3c). Class switch recombination (CSR) to IgG1 and IgG3 was found in three of four total patients, and the BCR sequences harbored 4-19 SHMs in the CDR1, CDR2, and framework regions. As CSR and SHM accumulation in BCR sequences are recognized as the hallmarks of chronic repeated exposure to antigens[18], it is possible that these clonotypes were developed by persistent exposure to autoantigens.

To investigate this, we synthesized four BCR heavy chain genes (Extended Table 4) to construct a single-chain variable fragment (scFv) phage display library, using the amplified light chain genes of AD patients (patient ID 6, 14, and 34) encoding the IGHV3-7/IGHJ1 VRLAEYFQN clonotype. We then performed a bio-panning of this library on the Aβ42 peptide, a well-known autoantigen in AD[16], and were able to find Aβ42 peptide-reactive scFv clones. When we expressed these clones as recombinant scFv-human Fc fusion proteins and performed an ELISA, two scFv clones with different kappa (κ) light chains (Fig. 3c, Extended figure 2a) showed reactivity against the Aβ42 peptide, with half-maximal effective concentrations (EC50) of 48.37 nM and 58.27 nM, respectively (Fig. 3d).

Our results support that the antibodies encoded by stereotypic AD patient-specific BCR clonotypes are developed by chronic exposure to autoantigens through the progression of disease.

Discussion

In this study, we found that AD patients share stereotypic BCR clonotypes not found in CG, and the degree of overlap between patient pairs were higher than that of CG pairs, despite the CG being exposed to the SARS-CoV2 spike antigen to develop a very high antibody titer and concerted BCR repertoire[15]. This observation strongly suggests the presence of a common immunogen among AD patients that coordinates the BCR repertoire into select directions. We also found that many of

these stereotypic AD patient-specific BCR clonotypes or their homologous clonotypes persist over a period of 14 and 53 weeks (Fig. 3a). In addition, when we examined this group of long-lasting homologous BCR clonotypes at the individual sequence level, we found not only CSR occurrences, but also the accumulation of SHM at a degree that was achievable only after the third injection of the COVID-19 vaccine (TP6) in the non-AD CG[15] (Fig. 3b). We also checked the CSR status and the level of SHM in the top 20 stereotypic AD patient-specific BCR clonotypes. One was selected for further testing of binding affinity to AD antigens, and of the 19 remaining clonotypes, 13 showed evidence of prior CSR events, and 16 showed evidence of prior SHMs of over 10 amino acid residues (Extended Table 5). These findings strongly claim that the exposure to this common immunogen was likely not a single event in most cases, but rather chronic. Collectively, our results strongly suggest the presence of common autoantigens that persist in AD patients and provoke a concerted and constitutive humoral immune response.

One of the drawbacks in our study is that the age of the CG was not matched with that of AD patients. BCR repertoires are known to be influenced by the aging process[19], and we have previously reported that many centenarians possess autoantibodies to the DNA-directed RNA polymerase II subunit RPB1, while being relatively rare in a younger CG[20]. Therefore, without an age and sex matched CG, it is difficult to rule out the possibility that the evolution of these stereotypic AD patient-specific BCRs are from the aging process rather than AD progression. In this regard, we tried to co-relate these stereotypic AD patient-specific BCR clonotypes with the pathogenic process of AD. Several autoimmune diseases were reported to show unique signatures on their BCR repertoires[21]. For example, the autoantibodies to the acetylcholine receptor and muscle-specific tyrosine kinase are found among most myasthenia gravis patients and very closely associated with pathogenicity[22]. In AD, over 70 autoantigens have been reported, to which autoantibodies were found in the blood or cerebrospinal fluid[3]. Among these autoantigens, Aβ has been recognized to play a pivotal role in AD pathogenesis. It causes synaptic impairment and neurodegeneration, thus contributing to the cognitive dysfunction observed in AD[23]. Two anti-Aβ antibodies, lecanemab and aducanumab, have been approved for clinical use in AD patients and few others are on clinical trials. For active immunotherapy, anti-Aβ vaccines capable of eliciting a polyclonal anti-Aβ antibody

response are also on clinical trials (NCT05462106 & NCT03461276). In this study, we focused on the IGHV3-7/IGHJ1 VRLAEYFQN clonotype and revealed that it encodes an antibody reactive to the Aβ42 peptide, a well-known autoantigen in AD[16]. With this finding, we believe that more stereotypic AD patient-specific BCR clonotypes may be related to the pathophysiology of AD.

Over the past decade, research findings highlighting the relationship between autoantibodies and the progression of brain diseases have provoked a paradigm shift in neurological diseases, opening opportunities for novel diagnostic modalities[24]. In this study, we identified 3,983 stereotypic AD patient-specific BCR clonotypes. It is also noteworthy that these clonotypes have been filtered by the strictest sequence contamination rationale reported so far, leaving little room for false positives. Blood-based biomarkers, like phosphorylated tau in plasma, hold great promise to revolutionize the diagnostic and prognostic work-up of AD in clinical practice[25]. BCR clonotypes are potential blood-based biomarkers and we have already reported the existence of disease-specific stereotypic BCR clonotypes among Sjogren syndrome patients[26] as well as COVID-19 patients[18]. Additionally, the accumulating single B cell sequencing data of AD patients can be combined with our findings to match the natural heavy and light chain pairs of BCR clonotypes to identify individual autoantigens, which could further facilitate the evaluation of BCR clonotypes as a blood-based biomarker. With these insights, we could also classify AD patients into subgroups based on the shared degree and types of stereotypic BCR clonotypes. A recent study reported that AD can be classified into three molecular subgroups by transcriptomics data, and the clinical features were found to differ in each group, suggesting the potential for personalized diagnosis and treatment[27].

In summary, we found that AD patients share many BCR clonotypes, of which homologous clonotypes are not or rarely found in healthy non-AD populations. By analyzing these commonly found stereotypic AD-patients specific clonotypes, we demonstrated that they were formed by repeated exposure to the antigens shared among the AD patients, such as Aβ. Additionally, as AD patients can be sub-grouped depending on the BCR signature, we believe that BCR repertoire analysis has potential to become a blood-based biomarker for the improved diagnosis and prognosis of AD.

Methods

Study subjects

We acquired data of all subjects, including demographic information, subtypes of dementia, Clinical Dementia Rating (CDR), comprehensive neuropsychological test results of Mini-Mental State Examination (MMSE), magnetic resonance imaging (MRI) results, and amyloid-PET results from the Korean Longitudinal Study on Cognitive Aging and Dementia (KLOSCAD)[28]. Each patient data was valid for 6 months except the amyloid-PET results. This study was approved under the recommendations of the Institutional Review Board (IRB) of the Seoul National University Bundang Hospital (SNUBH, IRB approval number, B-2102-664-305), South Korea. Protocols and manuals were also approved by the IRB of the SNUBH.

A total 55 of vaccinees who received the BNT162b2 mRNA vaccine were recruited, and their peripheral blood sampling was approved by the IRB of Seoul National University Hospital (SNUH, IRB approval number, 2102-032-1193). The details of the vaccination cohort data are described in a previous publication[15].

Blood sampling and peripheral blood mononuclear cell isolation

At least 15 ml of peripheral blood was collected per patient in EDTA tubes (BD Vacutainer K2EDTA tubes, 367525)[29]. Peripheral blood mononuclear cells (PBMCs) and plasma were separated by density-gradient centrifugation using Ficoll (Cytiva Ficoll-Paque Plus, 17-1440-02) solution. After the PBMC suspension was resuspended in CELLBANKER (Amsbio CELLBANKER 2, 11914), the collected PBMCs were immediately cryopreserved and stored in a Freezing container (Nalgene Mr. Frosty, C1562) for a week. The total RNA was isolated by TRIzol Reagent (Invitrogen TRIzol Reagents, 15596026) following the manufacturer's instructions.

Preparation of next-generation sequencing (NGS) library

For the NGS library preparation of BCR heavy chain sequences, genes encoding the variable region of the heavy chain (V_H) at the 5' forward region and the first constant region of the heavy chain (CH1) domain at the 3' reverse region were amplified by specific PCR primers as reported previously[18]. All primer sequences are listed in Extended Table 2. In short, 1 μg of total RNA was used as the input template for library preparation. Next, reverse transcription was performed using the SuperScript IV First-Strand Synthesis System (Invitrogen, 18091050) following the manufacturer's instructions with CH1 gene-specific primers for five immunoglobulin heavy chain isotypes containing UMI (unique molecular identifiers) barcodes. The UMI barcodes consisted of 12 random nucleotides, 14 base pairs in total; NNNNTNNNNTNNNN, and partial Illumina adapters (No. 1-8). After synthesis of the first-strand cDNA, it was purified with SPRI beads (Beckman Coulter AMPure XP Reagents, A63882) at a 1:1.8 ratio, and the second-strand cDNA was synthesized using a KAPA Biosystems kit (Roche KAPA HiFi Hotstart PCR kit, KK2502) with VH gene-specific primers (No. 9-14; 95°C for 3 min, 98°C for 30 s, 60°C for 45 s, and 72°C for 6 min). The double-stranded cDNA was purified through SPRI beads at a 1:1 ratio, then amplified using the KAPA Biosystems kit and double primers containing Illumina adapters and index sequences (No. 15, 16; 95°C for 3 min, 25 cycles of 98°C for 30 s, 60°C for 30 s, 72°C for 1 min, and 72°C for 5 min). The PCR products were then subjected to electrophoresis on a 1.5% agarose gel for separation by DNA lengths, and purified using the QIAquick gel extraction kit (QIAGEN, 28704) following the manufacturer's protocols. The final NGS libraries were obtained using SPRI beads at a 1:1 ratio and quantified with a 4200 TapeStation System (Agilent Technologies, G2991BA) for QC. Libraries with a single peak of appropriate length were subjected to NGS analysis using the Illumina Nocaseq6000 platform.

NGS data processing

The raw NGS forward (R1) and reverse (R2) reads were merged, filtered based on Phred scores, and subclustered using UMI with paired-end read merger (PEAR) v0.9.10 and Clustal Omega v1.2.4 with default settings[30,31]. The sequence

annotation process used the IMGT (International Immunogenetics Information System) constant gene database[32] to obtain the isotypes, and IgBLAST v1.17.1[33] to obtain VJ genes, CDR1/2/3 sequences, and the number of mutations from the corresponding V genes as previously described[18]. To exclude the possibility of contamination due to index swapping within Illumina flow cells between samples, reads with identical sequences including the UMI from different samples were removed on the basis that such an event is highly unlikely without contamination[34]. The combined NGS throughput of each sample is listed in Extended table 3. Sequencing reads that have survived through the entire process were considered to be productive BCR sequences.

BCR repertoire sharing and exclusion analysis

Shared BCR clonotypes are defined in this paper as the unique combination of a V gene, J gene, and CDR3 amino acid sequence present in 2 or more subjects within a cohort[18]. The shared BCR sequences in this context refer to the collective BCR sequences that can be grouped into each shared BCR clonotype. Clonotypes which consist only of naïve sequences or those with the IgM or IgD isotype with 1 or less somatic mutations were excluded from further analysis, in order to account only for shared clonotypes arising from shared antigenic stimuli[17]. Due to the potential risk of molecular contamination during the NGS library preparation stage, an additional strict *in silico* filtering step was performed. Shared clonotypes containing unique BCR nucleotide sequences specific to one patient that were completely eclipsed by another patient's BCR nucleotide sequences were considered as contamination, due to the low likelihood of 2 different individuals sharing the exact same somatic mutations at the nucleotide level.

The similarity scores based on shared cohort-specific BCR clonotypes were calculated as the cosine similarity of the two vectors each representing a subject's repertoire. The vectorization process involved utilizing one-hot encoding to create a vector with the length equal to the total number of cohort-specific and shared BCR clonotypes. The presence or absence of each clonotypes for the patient was encoded as either the value 1 or 0. The cosine similarity calculated as the inner

product divided by the product of the size of each vector can be interpreted as the number of shared cohort-specific shared clonotypes in a patient pair divided by the square root of the product of the total number of cohort-specific shared clonotypes for each patient.

Construction of scFv phage display library and bio-panning

For the V_H gene, the V3-7/J1 VRLAEYFQN clonotype was chosen for further testing of its affinity against the Aβ42 peptide, as this clonotype was the most shared among all AD patients. In total, 16 potential nucleotide sequences were included for analysis, and 9 out of 16 were observed at TP2. 4 out of the 9 identified sequences were synthesized in gBlocks Gene Fragments (Integrated DNA Technologies) depending on their somatic hypermutation status and sequence diversity of the CDR1, 2 regions at the amino acid level. For the construction of V_κ/V_λ shuffled libraries, the total RNA from patients who possessed the V3-7/J1 VRLAEYFQN clonotype at TP2 (AD9, AD14 and AD34) were used to synthesize cDNA using the SuperScript IV First-Strand Synthesis System (Invitrogen) with gene-specific primers amplifying the J_H and $C_{\kappa/\lambda}$ genes (No. 17, 18-20), respectively. Patient 35 also had the described heavy chain; however, experiments were preceded before they joined the second blood drawing. The cDNA was purified with SPRI beads (Beckman Coulter AMPure XP Reagents) at a 1:1.8 ratio, and the purified product was subjected to the first round of PCR synthesis using V_κ/V_λ and J_κ/J_λ gene-specific primers (No. 21-57; 95°C for 3 min, 4 cycles of 98°C for 1 min, 60°C for 1 min, 72°C for 1 min, and 72°C for 10 min). The PCR product underwent another purification step with SPRI beads at a 1:1 ratio, and the 4 synthesized V_H genes and V_κ/V_λ genes from the total RNA of 3 patients were amplified using a KAPA Biosystems kit (Roche) with following primers (No. 58-61; 95°C for 3 min, 27 cycles of 98°C for 30 s, 65°C for 30 s, 72°C for 1 min, and 72°C for 10 min). Then, the gene fragments of the V_H and V_κ/V_λ were each subjected to electrophoresis on a 1.5% agarose gel and purified via DNA size using the QIAquick gel extraction kit (QIAGEN) following the manufacturer's instructions. The 4 amplified V_H and V_κ/V_λ gene fragments from the 3 patients were mixed at an equal ratio to make a single pool of V_H and V_κ/V_λ genes, respectively. From the equally mixed pool of V_H and V_κ/V_λ , the same amounts (100

ng) of each gene fragment were subjected to overlap extension PCR to generate final scFv genes using a KAPA Biosystems kit (Roche) with overlap extension primers (No. 62, 63; 95°C for 3 min, 25 cycles of 98°C for 30 s, 65°C for 30 s, 72°C for 1 min 30 s, and 72°C for 10 min). The amplified scFv genes were then purified and cloned into a phagemid vector as described previously[35].

Phage display of the human scFv library with 1.1 x 10^9 colony-forming units was performed using 4 types of synthesized V_H gene fragments and shuffled V_κ/V_λ genes derived from the cDNA of patient No. 6, 14 and 34. The scFv library was subjected to four rounds of bio-panning against the human Aβ42 peptide (Tocris Bioscience Amyloid β-Peptide 1-42 human, 1428) as described previously[36,37]. Phage-display experiments including reamplification, rescue, and PEG precipitation of the library were performed with the protocols described in *Phage Display: A Laboratory Manual* [38]. Briefly, 2 μg of the Aβ42 peptide was coated to the surface of the immunotube with a high binding surface (SPL Life Sciences, 43015) for 16 hours at 4°C, and blocked with 5 ml of 3% bovine serum albumin (BSA) in PBS (w/v). Then the scFv phage library (about 10^{13} phages) was incubated for 2 hours at room temperature. During the first round of panning, the tube was washed once with 5 ml of 0.05% (v/v) Tween 20 (Sigma-Aldrich, P1379) in phosphate-buffered saline (PBST). After each round of panning, the phages bound to the coated antigen were eluted and rescued for the next round of bio-panning as described above. For the remaining rounds of panning, the number of washes was increased to three. To select antigen-reactive scFv phage clones, each phage clone from the titration plate of the last round of panning was subjected to phage ELISA, using the antigen-coated microtiter plates, as described previously[38]. Reactive scFv clones were analyzed through Sanger nucleotide sequencing using their phagemid DNA extracted from the individual phage clones using the same methods discussed previously[18]. A recombinant scFv protein tagged with human IgG1 FC (hFc) was expressed in the mammalian expression system and purified as described previously[18].

Enzyme-linked immunosorbent assay

The specificity and antigen affinity of individual phage scFv clones and recombinant

scFv-hFc fusion proteins were assessed by ELISA. 100 nM of human Aβ42 peptides dissolved in 50 μℓ of 0.1 M sodium bicarbonate coating buffer (pH 8.6) were added to each well of a 96-well microtiter plates (Corning, 3690) as described previously. In short, the plates were coated with the antigen by incubation at for 16 hours 4°C and blocked with 3% BSA in PBS (w/v) for 1 hour at 37°C. After shaking off the blocking solution, serially diluted phage supernatant (twofold) or scfv-hFc (fivefold, 8 dilutions, starting from 1 μM to 12.8 pM) in blocking buffer were added to the wells of the microtiter plates and incubated for 2 hours at 37°C. Then, the plates were washed three times with 0.05% PBST (v/v). For ELISAs with phage clones, the HRP-conjugated rabbit anti-human IgG Fc antibody (Invitrogen, 31423, 1:5,000) and the HRP-conjugated anti-M13 antibody (Sino Biological, 11973-MM05T-H, 1:4,000) were diluted in blocking buffer according to the manufacturer's protocols, then added to the wells and incubated for 1 hour at 37°C . This was performed in duplicate, to determine the amount of antigen-reactive antibody or M13 bacteriophage. In the ELISA for the recombinant scFv-hFc fusion proteins, the HRP-conjugated rabbit anti-human IgG Fc antibody (Invitrogen, 31423, 1:5,000) was added to the wells and incubated for 1 hour at 37°C. After three washes with 0.05% PBST, 2,2′-azino-bis-3-ethylbenzothiazoline-6-sulfonic acid solution (Invitrogen, 002024) was added to the wells as an HRP substrate. Absorbance was measured at 415 nm, respectively, using a microplate spectrophotometer (Thermo Fisher Scientific Inc., Multiskan GO). All assays were carried out in duplicate.

Statistical analysis

The P values were calculated using the student t-test unless otherwise specified during comparison of BCR repertoire characteristics. P-value < 0.05 was regarded as significant.

Data availability

The raw sequencing data and processed data have been submitted to the NCBI GEO database and is under review.

References

1. Selkoe, D.J. Alzheimer's disease: genes, proteins, and therapy. *Physiol Rev* **81**, 741-766 (2001).
2. Heneka, M.T.*, et al.* Neuroinflammation in Alzheimer's disease. *Lancet Neurol* **14**, 388-405 (2015).
3. Chatanaka, M.K., Sohaei, D., Diamandis, E.P. & Prassas, I. Beyond the amyloid hypothesis: how current research implicates autoimmunity in Alzheimer's disease pathogenesis. *Crit Rev Clin Lab Sci* **60**, 398-426 (2023).
4. Bettcher, B.M., Tansey, M.G., Dorothée, G. & Heneka, M.T. Peripheral and central immune system crosstalk in Alzheimer disease — a research prospectus. *Nature Reviews Neurology* **17**, 689-701 (2021).
5. Xu, H. & Jia, J. Single-Cell RNA Sequencing of Peripheral Blood Reveals Immune Cell Signatures in Alzheimer's Disease. *Front Immunol* **12**, 645666 (2021).
6. Huang, L.-T., Zhang, C.-P., Wang, Y.-B. & Wang, J.-H. Association of Peripheral Blood Cell Profile With Alzheimer's Disease: A Meta-Analysis. *Frontiers in Aging Neuroscience* **14**(2022).
7. Gate, D.*, et al.* Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. *Nature* **577**, 399-404 (2020).
8. Ciccocioppo, F.*, et al.* The Characterization of Regulatory T-Cell Profiles in Alzheimer's Disease and Multiple Sclerosis. *Sci Rep* **9**, 8788 (2019).
9. Larbi, A.*, et al.* Dramatic shifts in circulating CD4 but not CD8 T cell subsets in mild Alzheimer's disease. *J Alzheimers Dis* **17**, 91-103 (2009).
10. Pellicanò, M.*, et al.* Immune profiling of Alzheimer patients. *J Neuroimmunol* **242**, 52-59 (2012).
11. Xiong, L.L.*, et al.* Single-cell RNA sequencing reveals B cell-related molecular biomarkers for Alzheimer's disease. *Exp Mol Med* **53**, 1888-1901 (2021).
12. LeBien, T.W. & Tedder, T.F. B lymphocytes: how they develop and function. *Blood* **112**, 1570-1580 (2008).
13. 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).
14. Park, J.C.*, et al.* Association of B cell profile and receptor repertoire with the progression of Alzheimer's disease. *Cell Rep* **40**, 111391 (2022).
15. Park, S.*, et al.* An ancestral SARS-CoV-2 vaccine induces anti-Omicron variants antibodies by hypermutation. *bioRxiv*, 2023.2003.2015.532728 (2023).

16. Kronimus, Y., *et al.* Fc N-glycosylation of autoreactive Aβ antibodies as a blood-based biomarker for Alzheimer's disease. *Alzheimers Dement* (2023).
17. Ghraichy, M., *et al.* Maturation of the Human Immunoglobulin Heavy Chain Repertoire With Age. *Frontiers in Immunology* **11**(2020).
18. Kim, S.I., *et al.* Stereotypic neutralizing V(H) antibodies against SARS-CoV-2 spike protein receptor binding domain in patients with COVID-19 and healthy individuals. *Sci Transl Med* **13**(2021).
19. de Bourcy, C.F., *et al.* Phylogenetic analysis of the human antibody repertoire reveals quantitative signatures of immune senescence and aging. *Proc Natl Acad Sci U S A* **114**, 1105-1110 (2017).
20. Han, J., *et al.* Antibodies to the DNA-directed RNA polymerase II subunit RPB1 occur with highest frequency in centenarians. *Immun Ageing* **13**, 8 (2016).
21. Bashford-Rogers, R.J.M., *et al.* Analysis of the B cell receptor repertoire in six immune-mediated diseases. *Nature* **574**, 122-126 (2019).
22. Takamori, M. Myasthenia Gravis: From the Viewpoint of Pathogenicity Focusing on Acetylcholine Receptor Clustering, Trans-Synaptic Homeostasis and Synaptic Stability. *Front Mol Neurosci* **13**, 86 (2020).
23. Song, C., *et al.* Immunotherapy for Alzheimer's disease: targeting β-amyloid and beyond. *Transl Neurodegener* **11**, 18 (2022).
24. Prüss, H. Autoantibodies in neurological disease. *Nat Rev Immunol* **21**, 798-813 (2021).
25. Hansson, O., Blennow, K., Zetterberg, H. & Dage, J. Blood biomarkers for Alzheimer's disease in clinical practice and trials. *Nat Aging* **3**, 506-519 (2023).
26. Kim, M.K., *et al.* Sjögren's syndrome patients shares stereotypic B cell receptor clonotypes with distinct dysbiosis of the gut microbiome. *Investigative Ophthalmology & Visual Science* **64**, 3910-3910 (2023).
27. Li, H., *et al.* Identification of the molecular subgroups in Alzheimer's disease by transcriptomic data. *Front Neurol* **13**, 901179 (2022).
28. Han, J.W., *et al.* Overview of the Korean Longitudinal Study on Cognitive Aging and Dementia. *Psychiatry Investig* **15**, 767-774 (2018).
29. Lee, H., *et al.* Optimization of peripheral blood volume for in silico reconstitution of the human B-cell receptor repertoire. *FEBS Open Bio* **12**, 1634-1643 (2022).
30. Sievers, F. & Higgins, D.G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* **27**, 135-145 (2018).
31. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina

Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620 (2014).

32. Lefranc, M.P. IMGT databases, web resources and tools for immunoglobulin and T cell receptor sequence analysis, http://imgt.cines.fr. *Leukemia* **17**, 260-266 (2003).
33. Ye, J., Ma, N., Madden, T.L. & Ostell, J.M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* **41**, W34-40 (2013).
34. Griffiths, J.A., Richard, A.C., Bach, K., Lun, A.T.L. & Marioni, J.C. Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat Commun* **9**, 2667 (2018).
35. Andris-Widhopf, J., Steinberger, P., Fuller, R., Rader, C. & Barbas, C.F., 3rd. Generation of human scFv antibody libraries: PCR amplification and assembly of light- and heavy-chain coding sequences. *Cold Spring Harb Protoc* **2011**(2011).
36. Choi, J.R.*, et al.* BLI-Based Functional Assay in Phage Display Benefits the Development of a PD-L1-Targeting Therapeutic Antibody. *Viruses* **12**(2020).
37. Lee, Y., Kim, H. & Chung, J. An antibody reactive to the Gly63-Lys68 epitope of NT-proBNP exhibits O-glycosylation-independent binding. *Exp Mol Med* **46**, e114 (2014).
38. Barbas, C.F., Burton, D., Scott, J. & Silverman, G. Phage display. *A Laboratory Manual* (2001).

Acknowledgements

This research study was supported by the Korea Dementia Research Project grant through the Korea Dementia Research Center (KDRC), funded by the Ministry of Health & Welfare and Ministry of Science and ICT, Republic of Korea (grant number: HU20C0339). This work was supported by the BK21 FOUR program of the Education and Research Program for Future ICT Pioneers, Seoul National University in 2022. This work was supported by the Ministry of Science and ICT(MSIT) of the Republic of Korea and the National Research Foundation of Korea (NRF-2020R1A3B3079653). This work was supported by the Hyundai Car Chung Mong-Koo Foundation.

Additional information

Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to J. Chung.

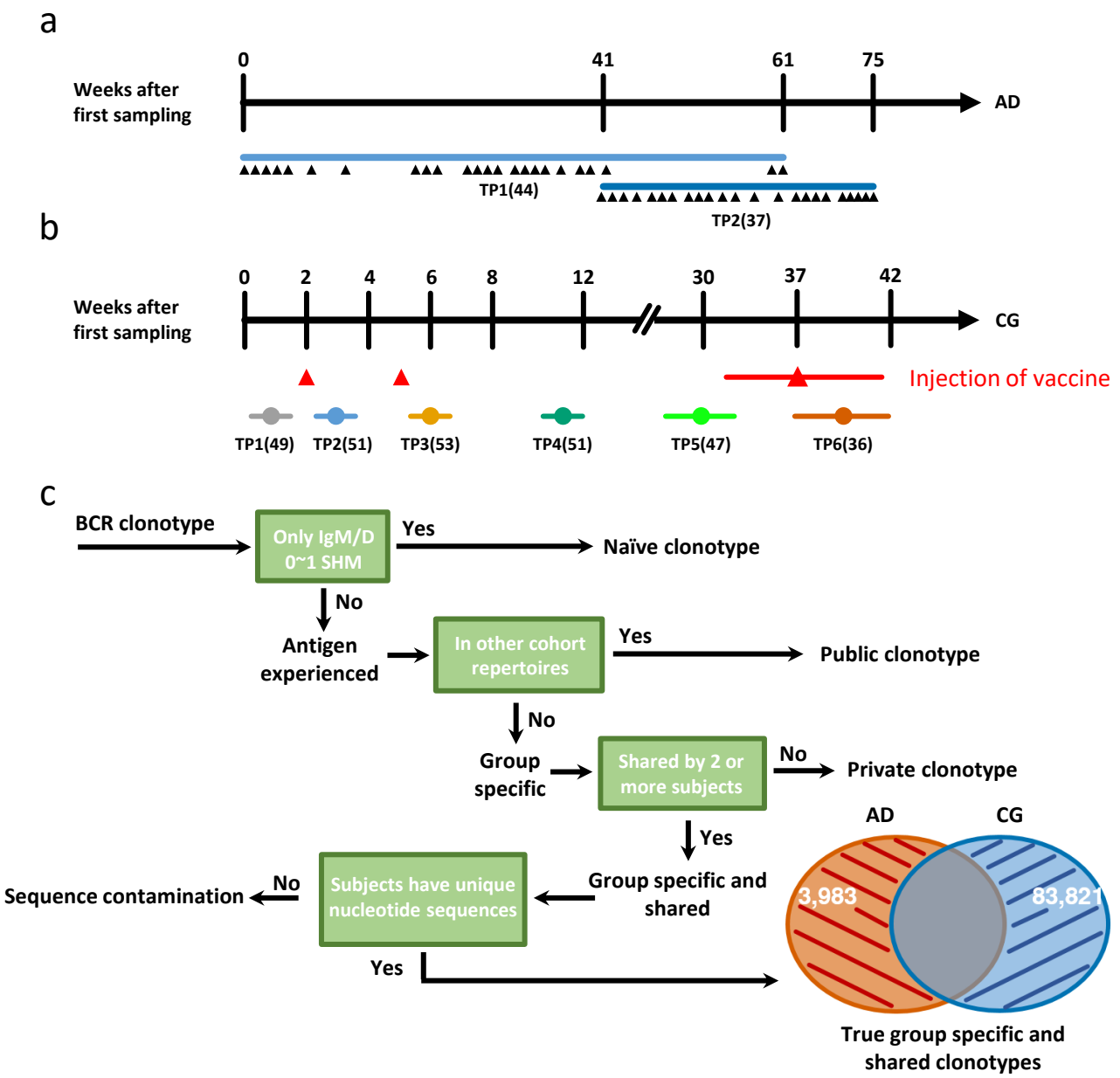


Figure 1. Outline of the sample acquisition schedule and BCR repertoire analysis pipeline. a, The peripheral blood samples of 44 patients from 2 different time points were collected for the Alzheimer's disease (AD) cohort. The blood samples collected at the second time point in this cohort are chronologically separated by 3.4 months minimum and 13.3 months maximum. The black arrowheads represent the exact sampling points of the samples in the given time frame. b, The vaccinated control group (CG) comprises of 55 vaccine recipients who were subjected to peripheral blood sampling before their first dose of vaccination (TP1), 1 week after the 1st dose (TP2), 1 week after the 2nd dose (TP3), 1 month after the 2nd dose (TP4), 6 months after the 2nd dose (TP5), and 1 month after the 3rd dose (TP6) c, To obtain BCR clonotypes that are representative of each cohort for further analyses, we performed sequence filtering based on the shown flow chart. Clonotypes which were comprised fully of naive sequences were removed and clonotypes unique to each cohort were obtained. Next, clonotypes that were present in two or more subjects within a cohort were selected. Due to the potential risk of sequence contamination during NGS library preparation and during sequencing, clonotypes that contained unique nucleotide sequences completely eclipsed by other patients were excluded.

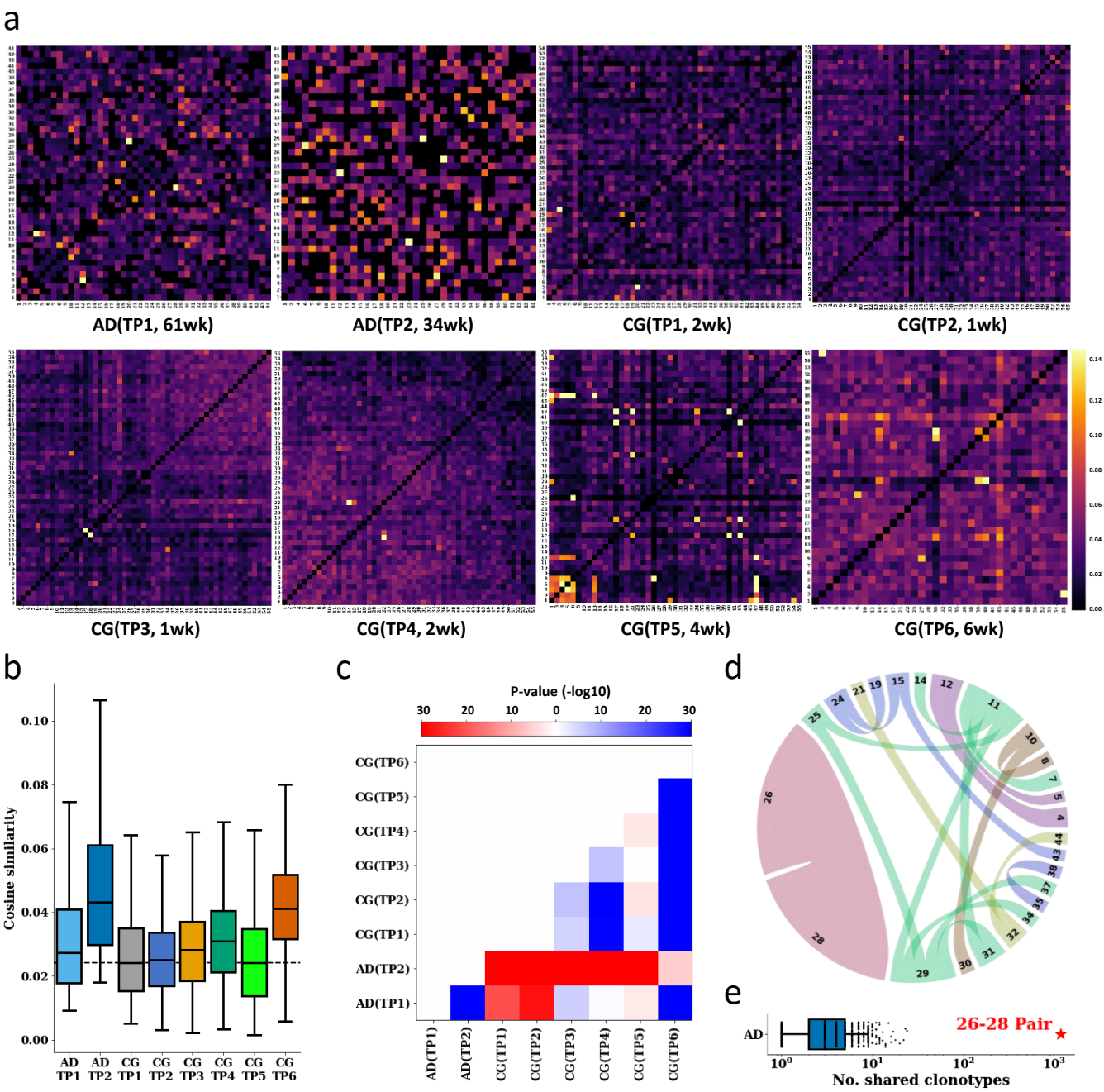
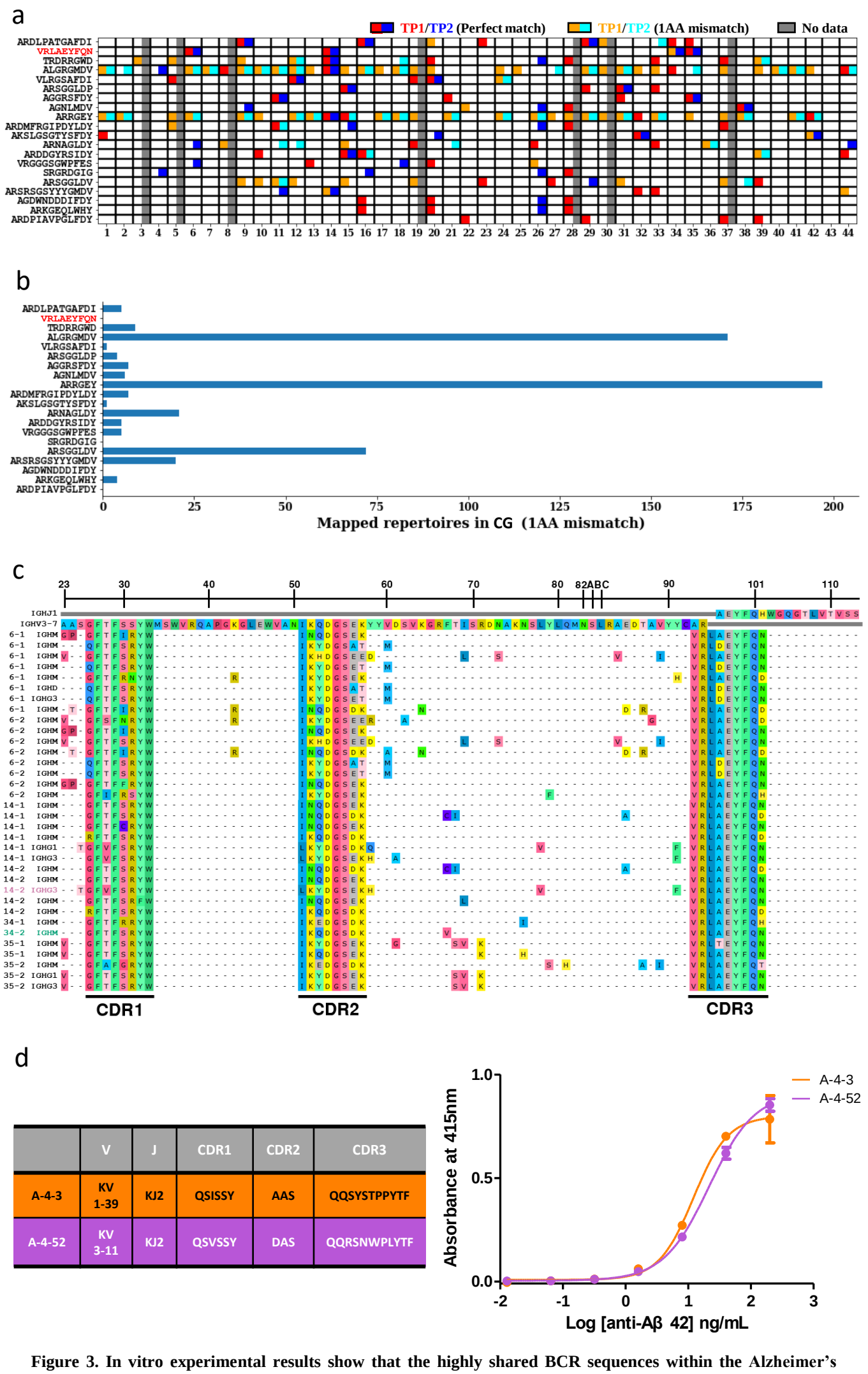
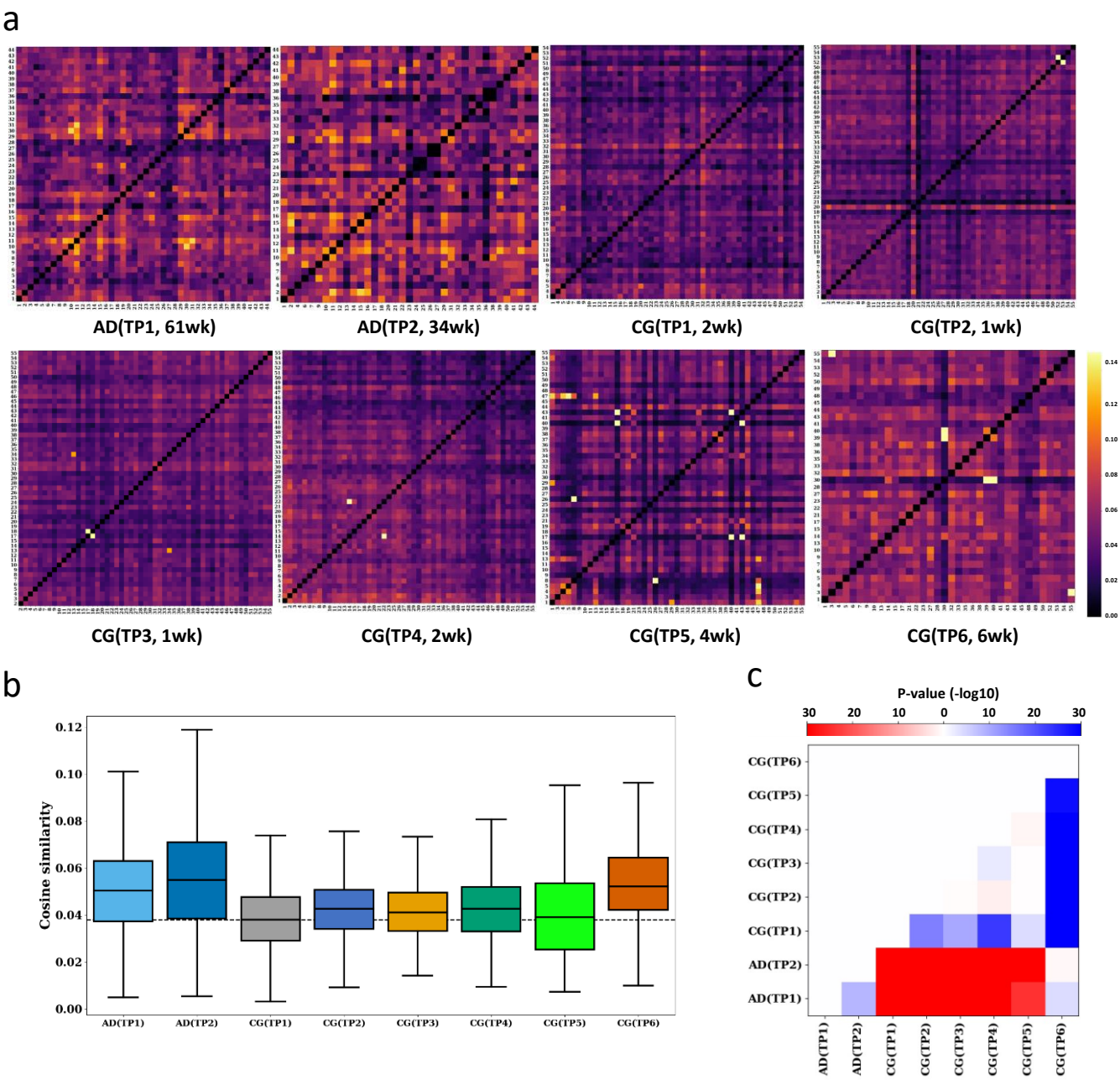


Figure 2. Alzheimer patients show distinctly high similarity within the AD cohort in the context of stereotypic and AD-specific BCR clonotypes. a, Heat map of the BCR similarity values between each subject pair within each group and time point. The similarity is calculated as the cosine similarity between two one-hot encoded linear vectors representing the presence (1) or absence (0) of group-specific and shared clonotypes within each subject's BCR repertoire. The peripheral blood collection time point is denoted inside parentheses. b, c, Boxplots representing the distribution of BCR similarity for all subject pairs within each group and a heat map representing the P-values between them. The dashed line represents the median value for the CG TP1 repertoires. The color of the heat maps indicate whether the median value of the group written on the y axis is larger (red) or smaller (blue), while the intensity of said colors is encoded by the P-value. d, Chord plot displaying the similarity matrix of the top 20 AD patient pairs. The similarities were calculated from the combined repertoires of both time points if applicable. The chord width represents the similarity value, with the thickest chord connecting patient 26 and 28 and representing the value 0.961, while the thinnest chord connecting patients 11 and 25 represents a value of 0.063. e, The number of shared clonotypes between each AD patient pair for the time point combined repertoires. The pair with the highest number of shared clonotypes is denoted with a red star.



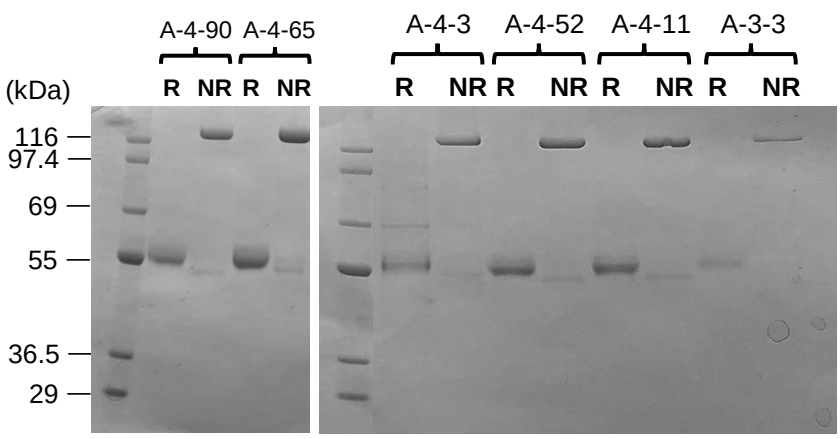
	V	J	CDR1	CDR2	CDR3
A-4-3	KV 1-39	KJ2	QSISSY	AAS	QQSYSTPPYTF
A-4-52	KV 3-11	KJ2	QSVSSY	DAS	QQRSNWPLYTF

Figure 3. In vitro experimental results show that the highly shared BCR sequences within the Alzheimer's disease patient cohort bind to the Aβ42 peptide. a, Colored table displaying the AD group specific BCR clonotypes that are shared by at least 4 different patients. The grey columns represent samples omitted from the TP2 stage due to patient withdrawal from study participation. Patients who possess similar clonotypes, or 1 amino acid mismatch in the CDR3 sequence, are also shown in the table, and coded with lighter colors. b, The number of CG repertoires that possess similar clonotypes for the AD specific clonotypes shown in a. c, Amino acid sequence alignment and representation of the highly shared and highly persistent BCR sequence. d, The light chain sequence information for the Aβ42 binders found via biopanning, A-4-3 and A-4-52. e, ELISA results of recombinant scFv A-4-3 and A-4-52 clones against Aβ42. The graph is normalized by the absorbance difference between Aβ42 and BSA (blocking only) signal.



Extended figure 1. Alzheimer's disease patients show distinctly high similarity within the AD cohort in the context of stereotypic BCR clonotypes, even without filtering by the control group. a, Heat map of the BCR similarity values between each subject pair within each group and time point. The similarity is calculated as the cosine similarity between two one-hot encoded linear vectors representing the presence (1) or absence (0) of shared and not group specific clonotypes within each subject's BCR repertoire. The peripheral blood collection time point is denoted inside parentheses. b, c, Boxplots representing the distribution of BCR similarity for all subject pairs of each group and the heat map representing the P-values between each group. The dashed line represents the median value for the CG TP1 repertoires. The color of the heat maps show whether the median value of the cohort written on the y axis is larger (red) or smaller (blue), while the intensity of said colors is encoded by the P-value.

a



Extended figure 2. Expression of scFv-hFc fusion proteins. A total of six out of eight Aβ42-binding scFv clones were successfully expressed in a mammalian expression system. All recombinant scFv proteins were tagged with human IgG1 FC (hFc). a. Representation of scFv-hFc fusion protein. The expected protein size is 56 kDa in the non-reducing state (NR) and double that in the reducing condition (R) because of the dimerization via disulfide bond with cysteine residues in the hFc tag.