



OPEN

Detection of *APP* gene recombinant in human blood plasma

Shigeki Mitsunaga^{1✉}, Naoko Fujito^{1,2}, Hirofumi Nakaoka³, Ryoko Imazeki⁴, Eiichiro Nagata⁴ & Ituro Inoue^{1✉}

The pathogenesis of Alzheimer's disease (AD) is believed to involve the accumulation of amyloid- β in the brain, which is produced by the sequential cleavage of amyloid precursor protein (APP) by β -secretase and γ -secretase. Recently, analysis of genomic DNA and mRNA from postmortem brain neurons has revealed intra-exonic recombinants of *APP* (gencDNA), which have been implicated in the accumulation of amyloid- β . In this study, we computationally analyzed publicly available sequence data (SRA) using probe sequences we constructed to screen *APP* gencDNAs. *APP* gencDNAs were detected in SRAs constructed from both genomic DNA and RNA obtained from the postmortem brain and in the SRA constructed from plasma cell-free mRNA (cf-mRNA). The SRA constructed from plasma cf-mRNA showed a significant difference in the number of *APP* gencDNA reads between SAD and NCI: the *p*-value from the Mann–Whitney *U* test was 5.14×10^{-6} . The transcripts were also found in circulating nucleic acids (CNA) from our plasma samples with NGS analysis. These data indicate that transcripts of *APP* gencDNA can be detected in blood plasma and suggest the possibility of using them as blood biomarkers for Alzheimer's disease.

Sporadic Alzheimer's disease (SAD) is the leading cause of dementia. It is characterized by amyloid- β (A β) accumulation in plaques and subsequent or partially preceding accumulation of abnormally phosphorylated tau in neurofibrillary tangles^{1,2}. Their accumulation causes neuronal damage, but the mechanism of pathogenesis is not fully understood. In addition, A β positron emission tomography (amyloid PET) has reported that A β accumulation begins as much as 20 years before the onset of dementia^{3,4}. This long-term latent accumulation makes it difficult to diagnose dementia and mild cognitive impairment (MCI), a pre-dementia stage, in the early clinical phases.

Accumulation of A β is examined by amyloid PET, which is radioactive, expensive, and available at limited facilities. The main components of A β , A β 42^{5,6}, phosphorylated tau, and total tau^{7,8} have also been measured in cerebrospinal fluid (CSF) as biomarkers for SAD, but the CSF collection remains invasive. Therefore, research and development of blood biomarkers for early diagnosis and screening tests are also underway^{9–13}. In addition to proteins/peptides measured in CSF, miRNAs^{14,15}, long non-coding RNAs (lncRNAs)^{16–18}, mRNA¹⁹, and circular RNA²⁰ have also been studied. Attempts to use these various blood RNAs as biomarkers generally involve combining many RNA species. They are, therefore, somewhat complex.

Brain atrophy is observed with age but to a marked degree in SAD^{21,22}. Atrophy of the amygdala, hippocampus, entorhinal cortex, and parahippocampal cortex is observed from early onset²³. Based on differences in the degree of atrophy in SAD and no cognitive impairment (NCI), the prediction of progression from MCI to SAD using MRI imaging has also been studied^{24–26}. The VSRAD, for example, analyzes regions of interest (ROI) in medial temporal structures such as the entorhinal cortex, hippocampus, and amygdala, where atrophy is more common in SAD patients.

On the other hand, genomic mosaicism due to somatic mutations exists in the brain and is observed in both SAD and NCI^{27,28}. Somatic mutations include single nucleotide variants (SNVs)^{29–32}, copy number variants (CNVs)^{33,34}, aneuploidy³⁵, and activation of retrotransposons^{36,37}. These genomic mosaicisms have been observed in postmortem brains and have been difficult to use as biomarkers. Recently, genomic cDNAs of *APP*

¹Laboratory of Human Genetics, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan. ²Department of Genetics, The Graduate University for Advanced Studies (SOKENDAI), Mishima 411-8540, Japan. ³Department of Cancer Genome Research, Sasaki Institute, Sasaki Foundation, Chiyoda-ku, Tokyo 101-0062, Japan. ⁴Department of Neurology, Tokai University School of Medicine, Isehara, Japan. ✉email: smitsunaga@nig.ac.jp; itinoue@nig.ac.jp

(*APP* gencDNA), one of the brain mosaicism, and their transcripts were reported in the postmortem brain³⁸, and another group has independently observed *APP* gencDNAs³¹.

APP gencDNA is formed by somatic recombination and integrated into the genomic DNA. It is characterized by the absence of introns and the presence of an intra-exonic junction. *APP* gencDNA is neuronal-specific, and the number of foci of *APP* gencDNA in DNA in situ hybridization (DISH) analysis is higher in SAD than in NCI. With age, an increase in DISH foci of *APP* gencDNA has also been reported using mouse models of SAD. In addition, the plasma nucleic acid levels have been reported to be increased in SAD compared to NCI³⁹. Because brain atrophy is caused by necrosis/apoptosis^{40,41}, these reports suggest that *APP* gencDNA and its transcripts are released into the plasma from injured neurons, as are other cfDNA and cfRNA. Since *APP* gencDNA is formed by recombination in somatic cells and not germline cells, its abundance is small, and it is considered that *APP* gencDNA in plasma is also very small. However, unlike proteins, nucleic acids are easily amplified, so it is possible that *APP* gencDNA could be detected even in plasma if it could be amplified. The concentration of high-mobility group box 1 (HMGB1) released from necrotic cells in CSF has also been higher in MCI than in SAD⁴⁰. Furthermore, the release of *APP* gencDNA from injured neurons may reflect the state of injured neurons in the brain more than other blood biomarkers.

These considerations suggest that early diagnosis of SAD, even before the onset of dementia, may be possible if *APP* gencDNA in nucleic acids released into plasma in response to neuronal damage can be detected. Therefore, we first analyzed publicly available sequence data (sequence read archive, SRA) to confirm the presence of *APP* gencDNA and its transcripts. For this purpose, we constructed probe sequences assuming homologous recombination and performed a computational screening using the probe sequence. Specifically, after confirming that the probe sequences worked in the SRA associated with the publication in which *APP* gencDNA was reported, we analyzed published SRAs constructed from genomic DNA and mRNAs; both were obtained from postmortem brains and confirmed the existence of *APP* gencDNA in the brain. In addition, we also confirmed that transcripts from *APP* gencDNA were present in the published SRA constructed from cell-free mRNA (cf-mRNA) in plasma. We further performed NGS analysis to detect *APP* gencDNA and its transcripts using our plasma samples' circulating nucleic acids (CNA).

Results

Computational screening of SRA associated with the first report of *APP* gencDNA

To confirm the presence of *APP* gencDNA and to get a complete picture of it, we designed probe sequences based on the *APP* mRNA sequence and screened publicly available sequence data (SRA) computationally. We created a total of 182,654 probe sequences assuming two-base homologous recombination (Fig. 1, Supplementary Table 1) (see “Construction of probe sequences and computational screening of SRAs” in “Materials and methods”). The constructed probe sequences were then used for computational screening of two runs, SRR7905478 and SRR7905479 (see “Analyzed SRAs in this study” in “Materials and methods”) of BioProject PRJNA493258, which were obtained by Pacbio-sequencing of the amplicon of the nested PCR of *APP* in postmortem human brain and associated with the first publication reporting the presence of *APP* gencDNA in postmortem human brain³⁸.

The probe sequences observed in SRR7905478 and SRR7905479 were shown in Supplementary Table 2. Thirty-eight probe sequences were positive, and various intra-exonic junction sequences were detected. Three probe sequences, that is, three recombination sites, out of 38 probe sequences detected were common with the recombination sites reported in the first publication of *APP* gencDNA³⁸. These results indicated that constructed probe sequences worked well for screening intra-exonic recombinant. The screening results are summarized in Table 1. The number of probe sequence positive reads for SRR7905479 in SAD cases was 190,934 out of 254,351 total reads; for SRR7905478 in NCI cases, it was 82,346 out of 360,290 total reads. Since these SRAs were constructed after nested PCR of *APP* amplifying between exons 1 and 18, general normalization using housekeeping genes was not possible, and the total read count normalized these read counts. The result was 0.751 for AD cases

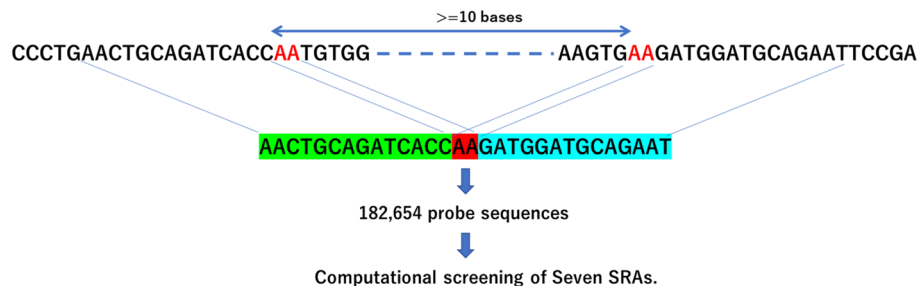


Figure 1. Construction of probe sequences for computational *APP* gencDNA screening. We searched for the same sequence of 2 bases at the 3' end and 2 bases at the 5' end. A probe sequence of 30 bases was created by combining 16 bases, including 2 homologous bases at the 3' end and 16 bases including 2 homologous bases at the 5' end, and removing one pair of 2 homologous bases. The probe sequence's construction region was the *APP* gene's coding region, and homologous regions at least 10 bases apart were selected. Probe sequences with identical sequences in the *APP* mRNA (*APP* transcript variant 1, Accession number NM_000484) and duplicates were removed, resulting in a final set of 182,654 probe sequences.

BioProject	PRJNA493258	
Run	SRR7905479	SRR7905478
Tissue/sample	Neuron/genomic DNA	
Sequencing	Pacbio	
Enrichment	Nested PCR	
Disease	SAD	NCI
Number of cases*	5	5
Average total read/case	50,870	72,058
<i>APP</i> gencDNA**	38,187	16,469
A β -producing*	3334	53

Table 1. Detection of *APP* gencDNA in the SRA of PRJNA493258. *Mixture of 5 cases. **Read count/case.

and 0.229 for NCI cases. These results indicate that *APP* intra-exonic recombination in SAD cases occurs more frequently than in NCI cases and are consistent with the previous reports³⁸.

Computational screening of SRAs constructed from genomic DNA and RNA of postmortem brains and plasma cf-mRNA

Having confirmed that the constructed probe sequences could detect *APP* gencDNAs, we next analyzed six SRAs constructed from genomic DNA or mRNA obtained from postmortem brains (Table 2). In SRAs constructed from genomic DNA using exon capture instead of nested PCR, *APP* gencDNA was hardly detected. However, *APP* gencDNA was indeed detected in SRAs constructed from mRNA. These results indicate that *APP* gencDNAs are certainly present and are transcribed in the brain, although their abundance is low.

APP gencDNAs and their transcripts in the brain should be released extracellularly with apoptosis/necrosis, the cause of brain atrophy. And released those should emerge in peripheral blood. Therefore, we next analyzed SRA (PRJNA574438) constructed from cf-mRNA (CNA) in blood plasma (Table 3). Three hundred thirty-one probe sequences in total were detected in PRJNA574438 (Supplementary Table 3), two identical to those observed in SRR7905480 of BioProject PRJNA493258. Many probe sequence reads were detected in SRR7905478

BioProject	PRJNA493258	PRJNA558504	PRJNA532465		PRJNA839035	PRJNA232669	PRJNA644383			
Run	SRR7905480	SRR9899152~4	SRR8898252~301							
Tissue/sample	Brain/genomic DNA				Brain/mRNA					
Sequencing	Illumina									
Enrichment	Capture-hybridization									
Disease	SAD	SAD	SAD	NCI	SAD	NCI	SAD	NCI	SAD	NCI
Number of cases	3	3	22	3	39	8	9	8	11	18
Average total reads/case	25,183,428	242,603,526	461,546,604	450,635,293	56,828,533	53,986,793	121,492,784	116,948,904	37,708,601	37,996,326
<i>APP</i> gencDNA*	2	10	0	0	2	2	11	12	6	4
A β -producing*	0	0	0	0	1	1	8	8	4	2

Table 2. Detection of *APP* gencDNA in SRAs constructed from genomic DNA or mRNA of postmortem brains. *Read count/case.

BioProject	PRJNA574438	
Run	SRR10192165~502	
Tissue/sample	plasma/cf-mRNA	
Sequencing	Illumina	
Enrichment	Capture-hybridization	
Disease	SAD	NCI
Number of cases	127	115
Average total read/case	22,777,062	17,823,100
<i>APP</i> gencDNA*	54	29
A β -producing*	52	29

Table 3. Detection of *APP* gencDNA in SRA constructed from cf-mRNA in blood plasma. *Read count/case.

and SRR7905479, constructed from nested PCR amplicons and associated with the publication in which APP gencDNA was first reported. Still, no positive probe sequences were common with them. In contrast, in the SRA constructed from postmortem brain mRNA, many positive probes were shared by SAD and NCI (Supplementary Table 3): the number of cases with positive probe sequences in PRJNA574438 was 125 of 127 for SAD and 96 of 115 for NCI.

Comparison of the number of APP gencDNA reads in an SRA from plasma cf-mRNA

When each read count of APP gencDNA read was normalized by dividing the read count of the housekeeping gene *GAPDH*, significant differences were observed between SAD and NCI: p -value by the Mann–Whitney U test was 5.14×10^{-6} (Fig. 2a). The distribution of read counts for the top ten probes in the positive cases was shown in Supplementary Fig. 1. For A β translation, frameshift did not occur in 207 of the 331 probe sequences (Supplementary Table 3). These reads were considered A β producible, but except for the probe sequence proseqff178928 positive, the number of other probe sequence positive reads was minimal. Table 4 shows the top 10 APP gencDNA-positive read counts, and Supplementary Table 3 shows all reads. Focusing on A β -producible reads, the average read count was 53/case for SAD and 33/case for NCI. On the other hand, no correlation between APP gencDNA normalized by GAPDH and MMSE score could be detected: for example, the r^2 between L-APP and MMSE score was 0.0016. There was also no correlation between the number of A β -producing recombinants normalized by GAPDH and the MMSE score: r^2 was 0.0056.

The most frequent probe sequence was proseqff178928, constructed as a recombinant at two bases homology region at the end of exon 14 and at the end of exon 15, which accounted for about 89% of the probe sequence positive reads. Its sequence was found to be identical to the junction sequence of exons 14 and 16 of the APP mRNA lacking exon 15, which is one of the APP isoforms and named L-APP mRNA⁴². L-APP mRNA is expressed in microglia and astrocytes⁴²; it is not neuron-specific. Therefore, we conducted the Mann–Whitney U test by dividing probe sequence-positive APP gencDNA into proseqff178928-positive, L-APP, and the rest (Fig. 2b,c). Both groups showed significant differences between SAD and NCI: p -value 5.54×10^{-6} for L-APP and p -value for APP gencDNAs minus L-APP was 8.81×10^{-5} . In addition, the Mann–Whitney U test, on the groups dividing according to their ability to produce amyloid- β , still showed significant differences between SAD and NCI (Fig. 2d,e): p -value 6.19×10^{-6} for APP gencDNAs including L-APP and p -value 1.04×10^{-3} for excluding L-APP.

APP mRNA with exon 8 spliced out is neuron-specific⁴³. So, we compared the number of reads for the exon 7 and exon 9 junction sequences normalizing with *GAPDH* between SAD and NCI. The junction sequence of exons 7 and 9 is not included in the probe sequence we constructed because it does not contain a homologous region. The p -value was 2.73×10^{-3} , which is significant but not a very small p -value.

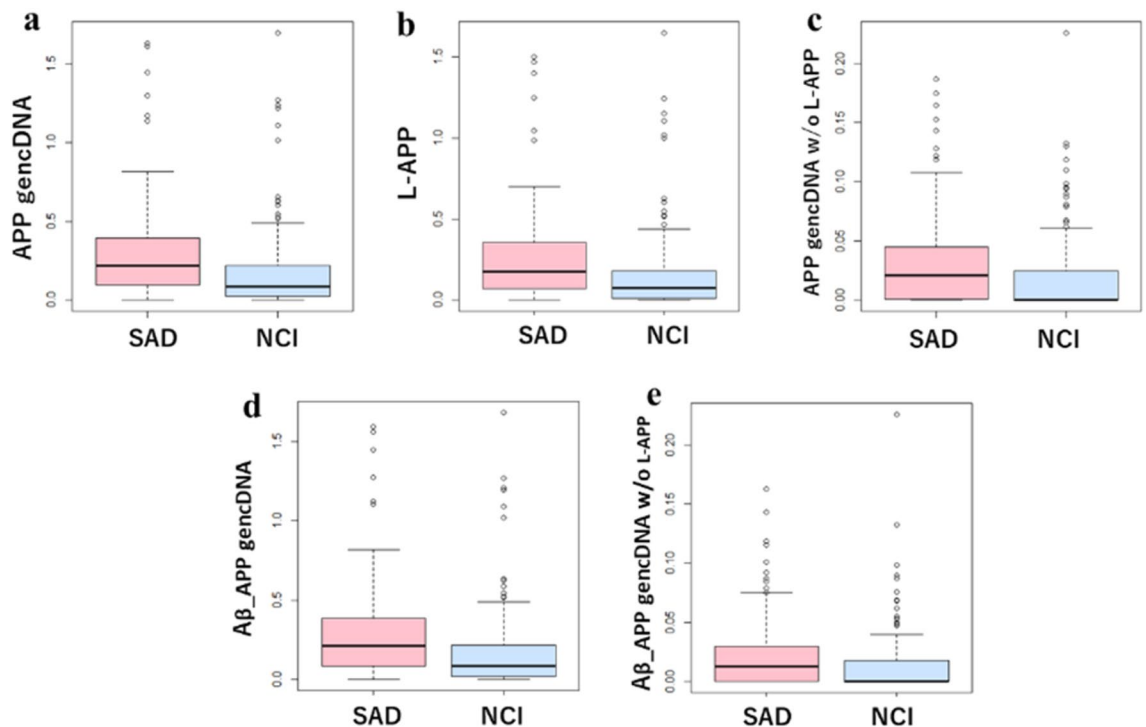


Figure 2. Comparison of APP gencDNA read counts in SAD and NCI. All read counts were normalized by GAPDH read counts. (a) Total read counts of APP gencDNA. (b) Read counts of L-APP. (c) Read counts of APP gencDNA minus L-APP. (d) Total read count of A β -producible APP gencDNA. (e) Total read count of A β -producible APP gencDNA minus L-APP. p -values by the Mann–Whitney U test are (a) 5.14×10^{-6} , (b) 5.54×10^{-6} , (c) 8.81×10^{-5} , (d) 6.19×10^{-6} , (e) 1.04×10^{-3} .

Probe ID	Probe sequence	Junction	AB-producible (no frameshift)	Number of positive case		Average read count per case	
				SAD	NCI	SAD	NCI
proseqff178928	AACACAGAAAACGAAGTTCTGGGTTGACA	ex14:ex15	○	124	91	49.7	30.5
proseqff97091	CTGAGGTGGAAGAAGAAGAACCCACAGAGA	ex6:ex6	○	21	11	0.4	0.2
proseqff93895	GAAGTAGCAGAGGAGGAAGAAGAAGGCC	ex6:ex6	○	17	6	0.2	0.1
proseqff90221	ATGCAGATGGGAGTGAAGAAGTGGCTGAGG	ex6:ex6	○	15	2	0.3	0.1
proseqff8330	CTCGGGCGCTGGAGGTACCTGAACTGC	ex2:ex3	○	9	4	0.1	0.1
proseqff95093	GAGGAGGAAGAAGTGGCTGAGGAACCTAC	ex6:ex6	○	11	5	0.1	0.1
proseqff184372	GAGCGCCACCTGTCCAGATGCAGAACTAGA	ex18:ex18		2	0	0.2	0.0
proseqff95821	GAAGAAGTGGCTGAGGAACCTACGAAGAA	ex6:ex6	○	9	3	0.1	0.0
proseqff153526	AGCCAACGAGAGACAGACAGCACACCCTAA	ex11:ex12		1	0	0.2	0.0
proseqff102427	ATGGTGATGAGGTAGAGGAACCTACGAAG	ex6:ex6	○	7	2	0.1	0.0

Table 4. Top 10 probe-sequence by read count in PRJNA574438.

NGS analysis of circulating nucleic acids in blood plasma and comparison of other SRAs

To confirm the presence and detectability of *APP* gencDNA in plasma, we purified CNA from our plasma samples and performed Nanopore-sequencing using PCR products amplified with primer set in exon 1 and exon 18 of the *APP* gene (Supplementary Table 5). Although *APP* gencDNA could not be detected in some samples, a variety of *APP* gencDNA was detected in many samples (Supplementary Table 6). This analysis using our plasma samples also detected seven identical probes to those detected in SRAs constructed from plasma cf-mRNA (CNA); 37 probe sequences were shared between the SRA constructed from plasma cf-mRNA and the SRAs constructed from mRNA from postmortem brain (Supplementary Table 4). Many probe sequences were commonly found, suggesting that *APP* gencDNA formation may not have occurred randomly.

Discussion

We could identify the reads with *APP* gencDNA sequences, that is, intra-exonic recombination sites, from several published sequence data (SRAs) using probe sequences constructed on the hypothesis of homologous recombination of two base duplications. Those are from the amplicon of a nested PCR of genomic DNA, enriched by capture hybridization of *APP* genomic DNA or mRNA from the postmortem brains. *APP* gencDNA sequences were detected in an SRA constructed from cf-mRNA in plasma, and we also detected them in circulating nucleic acid (CNA) in plasma by PCR amplification followed by Nanopore-sequencing using our plasma samples. Although the presence of *APP* gencDNA had been questioned⁴⁴, these results suggest that *APP* gencDNA formed by intra-exonic recombination is undoubtedly present, its quantity is meager, detection requires a high degree of amplification, and transcripts of *APP* gencDNA are detectable in plasma. In addition, none of the *APP* intra-exonic recombinants detected in this study were identical to human sequences in all 30 base lengths, except for proseqff8330, in BLAST analysis. The proseqff8330 was identical to *APP* transcript variant 5, 6, cDNA FLJ50491 (AK294534), and APP639 (lacking exon 2, 7, and 8, expressed in fetal tissue and liver). These results indicate that the *APP* intra-exonic recombinants detected in this study are not derived from other genes.

A comparison of the number of reads of *APP* gencDNA transcripts containing L-APP in plasma normalized by *GAPDH* showed a significant difference between SAD and NCI. *APP* gencDNA transcripts in plasma are derived from damaged brain cells; since the accumulation of A β begins as early as 20 years before the onset of dementia and A β itself is cytotoxic, the release of *APP* gencDNA transcripts may start at a relatively early clinical stage, that is the pre-clinical stage, of Alzheimer's disease. The report that the extent of YAP-dependent necrosis is more pronounced during the MCI phase than after the onset of Alzheimer's disease⁴⁰ supports this hypothesis. In summary, *APP* gencDNA transcripts, including L-APP in plasma, may serve as blood biomarkers for Alzheimer's disease and detect the early clinical or pre-clinical stages of Alzheimer's disease.

The Mann–Whitney *U* test showed some outliers. Changes over time in biomarkers associated with the development of dementia are not linear changes but changes represented by a sigmoid curve. Neuronal injury, which may be related to the release of nucleic acids, is also characterized by a sigmoid curve⁴⁵. That is the highest amount of released nucleic acids and YAP-dependent necrosis associated with the preclinical to early stages of Alzheimer's disease but not with the dementia phase. Therefore, outliers in the Mann–Whitney *U* test are likely to be immediately after the onset of dementia in SAD and before the start of dementia in NCI.

Computational screening of *PSEN1* and molecular chaperones^{46,47} reported to be involved in the pathogenesis of SAD was performed using probe sequences constructed in the same manner as *APP*. Intra-exonic recombination was observed in the transcripts of several heat shock protein genes (Supplementary Table 7). Still, the number of read counts per case was lower than for *APP* transcripts (*APP* gencDNA), and no difference was observed between SAD and NCI: for example, *HSP90AA1* normalized with *GAPDH* sequence with a *p*-value of 0.639 by the Mann–Whitney *U* test. These results indicate that intra-exonic recombination is not *APP*-specific and consistent with another report⁴⁸. Since reverse transcriptase activity is required for the formation of gencDNA³⁸, intra-exonic recombinants are likely to be abundant in the brain, where activation of transposable elements occurs⁴⁹.

The low number of reads for gencDNA compared to the *APP* gencDNA also suggests that the contribution of PCR-mediated recombination⁵⁰ to intra-exonic recombination of *APP* is minor.

In this study, proseqff178928, designed to detect sequences formed by homologous recombination between two bases at the 3' end of exon 14 and two bases at the 3' end of exon 15 of the *APP* gene, was the most frequently detected probe. And since the putative homologous recombination site detected by this probe is the same sequence as the junction site of exons 14 and 16 of L-*APP*, the *APP* isoform lacking exon 15, proseqff178928 could detect not only the *APP* gencDNA transcript but also L-*APP*. L-*APP* is expressed in astrocytes and microglia, which are involved in innate immunity in the brain, and these cells are also involved in removing A β plaques and aggregated tau proteins by activating and releasing inflammatory cytokines⁵¹. Since activated astrocytes and microglia cause programmed cell death (PCD) of them⁵¹, L-*APP* transcripts will be released from these dead cells. L-*APP* are expected to be more abundant than *APP* gencDNA because they are derived from normal genes, as opposed to transcripts derived from *APP* gene DNA, which are caused by somatic mutations. Therefore, the sequence that proseqff178928 was detecting was most likely derived primarily from L-*APP*, not from *APP* gencDNA. Astrocytes have been reported to upregulate glial fibrillary acidic protein (GFAP) when activated⁵². This GFAP is one of the blood biomarkers of Alzheimer's disease, and its blood levels are reduced by treatment with anti-A β antibodies⁵³. Therefore, L-*APP* detected by proseqff178928 could be used as a blood biomarker for Alzheimer's disease and as a biomarker for therapeutic efficacy since it is derived from astrocytes as well as GFAP and is detected more frequently in SAD than in NCI.

Materials and methods

Analyzed SRAs in this study

Publicly available seven sequence-read archives (SRAs) were analyzed in this study: PRJNA493258³⁸ (Tables 1 and 2), PRJNA558504⁵⁴, PRJNA532465³¹, PRJNA839035⁵⁵, PRJNA232669⁵⁶, PRJNA644383⁵⁷ (Table 2), and PRJNA574438¹⁹ (Table 3). SRR7905478 and SRR7905479 from PRJNA493258 were constructed by PacBio sequencing of the amplicon obtained from the nested PCR of *APP* using five normal human brains (SRR7905478) and five SAD patient brains (SRR7905479). SRR7905480 of PRJNA493258, PRJNA558504, and PRJNA532465 were constructed by exon capture hybridization and Illumina sequencing of genomic DNA obtained from the brain. PRJNA839035, PRJNA232669, and PRJNA644383 are the SRA of RNA-seq of brain mRNA. PRJNA574438 was constructed from cell-free messenger RNA (cf-mRNA) in the blood plasma of 127 SAD patients and 116 age-matched controls. There were 95 duplicates in this SRA, and one disease state was not identified. In the case of duplicates, the average read count was used in the analysis. The sequence data for which the disease state was not identified were excluded from the data analysis.

Construction of probe sequences and computational screening of SRAs

Probe sequences for screening of intra-exonic junction in *APP* cDNA were constructed based on the mRNA sequence (*APP* transcript variant 1, Accession number NM_000484) as follows: since homologous recombination occurs between regions of homology of two or more bases, 14 bases upstream of a two-base homologous region, a two-base homologous sequence, and 14 bases downstream of a homologous region were combined. This procedure resulted in a 30-base probe sequence (Fig. 1); only the coding sequence of *APP* was targeted, and the distance between homologous regions was at least ten bases to reduce interference from repeated sequences such as two-base and three-base sequences. Duplications were eliminated, and probe sequences with identical sequences in the *APP* mRNA (*APP* transcript variant 1, Accession number NM_000484) were deleted, resulting in a final set of 182,654 probe sequences (Supplementary Table 1). These probe sequences can detect intra-exonic junctions of not only 2-base homologous sequences but also longer than 2-base homologous sequences. Using this set of probe sequences, each fastq file was screened by the `bbduk` command (`sourceforge.net/projects/bbmap/`): `cat flistL.txt | while read I # flist: list of fastq file; do echo $I >> wc_SRA1; /home/bbmap/bbduk.sh -da in=${I} outm=clean.fq ref=reference.fa rename=t k=30; # reference.fa: multi fasta file of probes; grep -o @SRR clean.fq | wc -w >> wc_SRA1; cat clean.fq >> wc_SRA1; done`. The extracted fastq files with probe sequences were analyzed in EXCEL. R software and packages ("exactRankTests", "tidyverse" and "beeswarm") were used for statistical analysis and figure drawing. Reads in which two or more of the same probe sequence were detected were considered derived from PCR artifacts and excluded from the count. This was more common in SRR7905478 and SRR7905479, which were constructed from nested PCR amplicons.

SAD cases

All studies were conducted in accordance with the Declaration of Helsinki and after approval of the institutional review committees of two participating institutions: Institutional Review Board for Clinical Research, Tokai University (approval number 22R-180), and Ethical Review Committee on Medical and Biological Research Involving Human Subjects, National Institute of Genetics (approval number R2-14). After diagnosed to be SAD and obtaining written informed consent at the Tokai University School of Medicine, plasma specimens from 11 SAD patients were collected (Supplementary Table 8). All plasma specimens were collected with the RNA Complete BCT (Streck), and separated plasma was stored at -80 °C until use.

Sequencing analysis of circulating nucleic acid in blood plasma

Circulating nucleic acids were extracted from 1 mL of blood plasma using QIAamp Circulating Nucleic Acid Kit (QIAGEN) and eluted in 19 μ L. Double-stranded cDNA was synthesized using 16 μ L of eluate and LunaScript RT SuperMix Kit (NEB) in a 20 μ L total reaction mixture. After PCR amplification (Supplementary Table 5) using all the cDNA solutions and Q5Hot Start High-Fidelity 2 \times Master Mix, the sequencing library was prepared with Native barcoding following ONT instructions. Nanopore-sequencing was performed using 30 ng library

solution in 12 μ L obtained from 16 plasma samples. The *APP* gencDNA sequences were extracted using probe sequences described above.

Data availability

The datasets generated and analysed during the current study are available with links to BioProject accession number PRJDB16023 in the DDBJ BioProject database.

Received: 1 June 2023; Accepted: 2 December 2023

Published online: 07 December 2023

References

- Jagust, W. Imaging the evolution and pathophysiology of Alzheimer disease. *Nat. Rev. Neurosci.* **19**, 687–700 (2018).
- Lee, W. J. *et al.* Regional A β -tau interactions promote onset and acceleration of Alzheimer's disease tau spreading. *Neuron* **110**, 1932–1943e5 (2022).
- Villemagne, V. L. *et al.* Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: A prospective cohort study. *Lancet Neurol.* **12**, 357–367 (2013).
- Villeneuve, S. *et al.* Existing Pittsburgh Compound-B positron emission tomography thresholds are too high: Statistical and pathological evaluation. *Brain* **138**, 2020–2033 (2015).
- Schirinzi, T. *et al.* Levels of amyloid-beta-42 and CSF pressure are directly related in patients with Alzheimer's disease. *J. Neural Transm. (Vienna)* **124**, 1621–1625 (2017).
- Hansson, O. *et al.* CSF biomarkers of Alzheimer's disease concord with amyloid- β PET and predict clinical progression: A study of fully automated immunoassays in BioFINDER and ADNI cohorts. *Alzheimers Dement.* **14**, 1470–1481 (2018).
- Petrice, M. *et al.* CSF phosphorylated tau as an indicator of subsequent tau accumulation. *Neurobiol. Aging* **117**, 189–200 (2022).
- Horie, K., Barthélemy, N. R., Sato, C. & Bateman, R. J. CSF tau microtubule binding region identifies tau tangle and clinical stages of Alzheimer's disease. *Brain* **144**, 515–527 (2021).
- Schindler, S. E. *et al.* High-precision plasma β -amyloid 42/40 predicts current and future brain amyloidosis. *Neurology* **93**, e1647–e1659 (2019).
- Karikari, T. K. *et al.* Blood phosphorylated tau 181 as a biomarker for Alzheimer's disease: A diagnostic performance and prediction modelling study using data from four prospective cohorts. *Lancet Neurol.* **19**, 422–433 (2020).
- Mattsson-Carlgen, N. *et al.* Longitudinal plasma p-tau217 is increased in early stages of Alzheimer's disease. *Brain* **143**, 3234–3241 (2020).
- Nakamura, A. *et al.* High performance plasma amyloid- β biomarkers for Alzheimer's disease. *Nature* **554**, 249–254 (2018).
- Jia, L. *et al.* Blood neuro-exosomal synaptic proteins predict Alzheimer's disease at the asymptomatic stage. *Alzheimers Dement.* **17**, 49–60 (2021).
- Shigemizu, D. *et al.* Risk prediction models for dementia constructed by supervised principal component analysis using miRNA expression data. *Commun. Biol.* **2**, 77 (2019).
- Abdollahi, S., Zare-Chahoki, A., Noorbakhsh, F. & Gorji, A. A review of molecular interplay between neurotrophins and miRNAs in neuropsychological disorders. *Mol. Neurobiol.* <https://doi.org/10.1007/s12035-022-02966-5> (2022).
- Fotuhi, S. N., Khalaj-Kondori, M., Feizi, M. A. H. & Talebi, M. Long non-coding RNA BACE1-AS may serve as an Alzheimer's disease blood-based biomarker. *J. Mol. Neurosci.* **69**, 351–359 (2019).
- Liu, Y. *et al.* Non-coding RNAs as novel regulators of neuroinflammation in Alzheimer's disease. *Front. Immunol.* **13**, 908076 (2022).
- Blount, G. S., Coursey, L. & Kocerha, J. MicroRNA networks in cognition and dementia. *Cells* **11**, 1882 (2022).
- Toden, S. *et al.* Noninvasive characterization of Alzheimer's disease by circulating, cell-free messenger RNA next-generation sequencing. *Sci. Adv.* **6**, eabb1654 (2020).
- Zheng, D. *et al.* Screening of human circular RNAs as biomarkers for early onset detection of Alzheimer's disease. *Front. Neurosci.* **16**, 878287 (2022).
- Guo, H. *et al.* Evaluation of common structural brain changes in aging and Alzheimer disease with the use of an MRI-based brain atrophy and lesion index: A comparison between T1WI and T2WI at 1.5T and 3T. *Am. J. Neuroradiol.* **35**, 504–512 (2014).
- Josephs, K. A. *et al.* Protein contributions to brain atrophy acceleration in Alzheimer's disease and primary age-related tauopathy. *Brain* **143**, 3463–3476 (2020).
- Planche, V. *et al.* Structural progression of Alzheimer's disease over decades: The MRI staging scheme. *Brain Commun.* **4**, fcac109 (2022).
- Syaifullah, A. H. *et al.* Machine learning for diagnosis of AD and prediction of MCI progression from brain MRI using brain anatomical analysis using diffeomorphic deformation. *Front. Neurol.* **11**, 576029 (2021).
- Dutt, S., Yanrong, L., Mather, M., Nation, D. A., Alzheimer's Disease Neuroimaging Initiative. Brainstem substructures and cognition in prodromal Alzheimer's disease. *Brain Imaging Behav.* **15**, 2572–2582 (2021).
- Matsuda, H. MRI morphometry in Alzheimer's disease. *Ageing Res. Rev.* **30**, 17–24 (2016).
- Rabbito, A., Dulewicz, M., Kulczyńska-Przybyk, A. & Mroczko, B. Biochemical markers in Alzheimer's disease. *Int. J. Mol. Sci.* **21**, 1989 (2020).
- Lodato, M. A. *et al.* Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555–559 (2018).
- Costantino, I., Nicodemus, J. & Chun, J. Genomic mosaicism formed by somatic variation in the aging and diseased brain. *Genes (Basel)* **12**, 1071 (2021).
- Abascal, F. *et al.* Somatic mutation landscapes at single-molecule resolution. *Nature* **593**, 405–410 (2021).
- Park, J. S. *et al.* Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nat. Commun.* **10**, 3090 (2019).
- Miller, M. B. *et al.* Somatic genomic changes in single Alzheimer's disease neurons. *Nature* **604**, 714–722 (2022).
- Christer, W. D. *et al.* Neurons with complex karyotypes are rare in aged human neocortex. *Cell Rep.* **26**, 825–835.e7 (2019).
- Lee, W. P. *et al.* Copy number variation identification on 3,800 Alzheimer's disease whole genome sequencing data from the Alzheimer's Disease Sequencing Project. *Front. Genet.* **12**, 752390 (2021).
- Kingsbury, M. A., Yung, Y. C., Peterson, S. E., Westra, J. W. & Chun, J. Aneuploidy in the normal and diseased brain. *Cell. Mol. Life Sci.* **63**, 2626–2641 (2006).
- Richardson, S. R., Morell, S. & Faulkner, G. J. L1 retrotransposons and somatic mosaicism in the brain. *Annu. Rev. Genet.* **48**, 1–27 (2014).
- Guo, C. *et al.* Tau activates transposable elements in Alzheimer's disease. *Cell Rep.* **23**, 2874–2880 (2018).
- Lee, M. H. *et al.* Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* **563**, 639–645 (2018).

39. Pai, M. C., Kuo, Y. M., Wang, I. F., Chiang, P. M. & Tsai, K. J. The role of methylated circulating nucleic acids as a potential biomarker in Alzheimer's disease. *Mol. Neurobiol.* **56**, 2440–2449 (2019).
40. Tanaka, H. *et al.* YAP-dependent necrosis occurs in early stages of Alzheimer's disease and regulates mouse model pathology. *Nat. Commun.* **11**, 507 (2020).
41. Erten-Lyons, D., Jacobson, A., Kramer, P., Grupe, A. & Kaye, J. The FAS gene, brain volume, and disease progression in Alzheimer's disease. *Alzheimers Dement.* **6**, 118–124 (2010).
42. Sandbrink, R., Banati, R., Masters, C. L., Beyreuther, K. & König, G. Expression of L-APP mRNA in brain cells. *Ann. N. Y. Acad. Sci.* **695**, 183–189 (1993).
43. Dawkins, E. & Small, D. H. Insights into the physiological function of the β -amyloid precursor protein: Beyond Alzheimer's disease. *J. Neurochem.* **129**, 756–769 (2014).
44. Kim, J. *et al.* APP gene copy number changes reflect exogenous contamination. *Nature* **584**, E20–E28 (2020).
45. Jack, C. R. Jr. *et al.* Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* **9**, 119–128 (2010).
46. Gorantla, N. V. & Chinnathambi, S. Tau Protein squired by molecular chaperones during Alzheimer's disease. *J. Mol. Neurosci.* **66**, 356–368 (2018).
47. Campanella, C. *et al.* Heat shock proteins in Alzheimer's disease: Role and targeting. *Int. J. Mol. Sci.* **19**, 2603 (2018).
48. Palmer, C. R., Liu, C. S., Romanow, W. J., Lee, M. H. & Chun, J. Altered cell and RNA isoform diversity in aging Down syndrome brains. *Proc. Natl. Acad. Sci. USA* **118**, e2114326118 (2021).
49. Shpyleva, S., Melnyk, S., Pavliv, O., Pogribny, I. & James, S. J. Overexpression of LINE-1 retrotransposons in autism brain. *Mol. Neurobiol.* **55**, 1740–1749 (2018).
50. Shafikhani, S. Factors affecting PCR-mediated recombination. *Environ. Microbiol.* **4**, 482–486 (2002).
51. Rajesh, Y. & Kanneganti, T. D. Innate immune cell death in neuroinflammation and Alzheimer's disease. *Cells* **11**, 1885 (2022).
52. Hol, E. M. & Pekny, M. Glial fibrillary acidic protein (GFAP) and the astrocyte intermediate filament system in diseases of the central nervous system. *Curr. Opin. Cell Biol.* **32**, 121–130 (2015).
53. Pontecorvo, M. J. *et al.* Association of donanemab treatment with exploratory plasma biomarkers in early symptomatic Alzheimer disease: A secondary analysis of the TRAILBLAZER-ALZ randomized clinical trial. *JAMA Neurol.* **79**, 1250–1259 (2022).
54. Lee, M. H. *et al.* Reply to: APP gene copy number changes reflect exogenous contamination. *Nature* **584**, E29–E33 (2020).
55. Caldwell, A. B. *et al.* Transcriptomic profiling of sporadic Alzheimer's disease patients. *Mol. Brain* **15**, 83–89 (2022).
56. Scheckel, C. *et al.* Regulatory consequences of neuronal ELAV-like protein binding to coding and non-coding RNAs in human brain. *Elife* **5**, e10421 (2016).
57. Nativio, R. *et al.* An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. *Nat. Genet.* **52**, 1024–1035 (2020).

Acknowledgements

We thank Drs. S. Toden and A. Ibarra (Molecular Stethoscope Inc.) and J. Chun (Sanford Burnham Prebys Medical Discovery Institute) for providing valuable information on SRAs; Dr. A. K. Munirajan (University of Madras) for helpful discussion; Drs. R. Sugimoto and P.T. Nguyen for advising on informatics (National Institute of Genetics); Y. Sato, J. Kitayama, J. Kajiwara, and A. Oishi for technical assistance (National Institute of Genetics).

Author contributions

S.M. conceived the project. I.I. developed a joint research system. S.M. and H.N. designed experiments. S.M. and N.F. analyzed SRAs. S.M. analyzed experimental data and wrote the manuscript. N.F., H.N., E.N., R.I., and I.I. critically reviewed and edited the manuscript. E.N. and R.I. collected SAD samples.

Competing interests

Research Organization of Information and Systems has filed the following patent application on the subject matter of this publication: Patent application number 2021-183638PCT entitled “Diagnostic marker detection methods and diagnostic kit,” filed on Nov 10, 2021.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-48993-7>.

Correspondence and requests for materials should be addressed to S.M. or I.I.

Reprints and permissions information is available at 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023