Whole exome sequence-based association analyses of plasma amyloid-β in African and European Americans; the Atherosclerosis Risk in Communities-Neurocognitive Study

Jeannette Simino, Zhiying Wang, Jan Bressler, Vincent Chouraki, Qiong Yang, Steven G. Younkin, Sudha Seshadri, Myriam Fornage, Eric Boerwinkle, Thomas H. Mosley, Jr.

1 Gertrude C. Ford MIND Center, University of Mississippi Medical Center, Jackson, Mississippi, United States of America. 2 Department of Data Science, John D. Bower School of Population Health, University of Mississippi Medical Center, Jackson, Mississippi, United States of America. 3 Human Genetics Center, Department of Epidemiology, Human Genetics & Environmental Sciences, School of Public Health, University of Texas Health Science Center, Houston, Texas, United States of America. 4 Lille University, Inserm, CHU Lille, Institut Pasteur de Lille, U1167 - RID-AGE - Risk factors and molecular determinants of aging-related diseases; Lille, France. 5 Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America. 6 The National Heart, Lung, and Blood Institute’s Framingham Heart Study, Framingham, Massachusetts, United States of America. 7 Department of Neuroscience, Mayo Clinic College of Medicine, Mayo Clinic Jacksonville, Jacksonville, Florida, United States of America. 8 Department of Neurology, University of Minnesota, Minneapolis, Minnesota, United States of America. 9 The Brown Foundation Institute of Molecular Medicine, Research Center for Human Genetics, The University of Texas Health Science Center, Houston, Texas, United States of America. 10 Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, United States of America. 11 Department of Medicine, University of Mississippi Medical Center, Jackson, Massachusetts, United States of America.

*jsimino@umc.edu

Abstract

Objective

We performed single-variant and gene-based association analyses of plasma amyloid-β (aβ) concentrations using whole exome sequence from 1,414 African and European Americans. Our goal was to identify genes that influence plasma aβ42 concentrations and aβ42:aβ40 ratios in late middle age (mean = 59 years), old age (mean = 77 years), or change over time (mean = 18 years).

Methods

Plasma aβ measures were linearly regressed onto age, gender, APOE ε4 carrier status, and time elapsed between visits (fold-changes only) separately by race. Following inverse normal transformation of the residuals, seqMeta was used to conduct race-specific single-variant and gene-based association tests while adjusting for population structure. Linear regression models were fit on autosomal variants with minor allele frequencies (MAF) ≥ 1%. T5 burden and Sequence Kernel Association (SKAT) gene-based tests assessed functional interactions between genes and aβ levels.
well-established procedures to request this data from the ARIC Coordinating Center at the University of North Carolina—Chapel Hill. Instructions for submitting a data request can be found at: http://www2.csc.cnc.unc.edu/aric/distribution-agreements.

Funding: The ARIC Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). ARIC Neurocognitive data was collected through the joint support of the National Heart, Lung, and Blood Institute; National Institute of Neurological Disorders and Stroke; National Institute on Aging; and the National Institute of Deafness and Other Communication Disorders (U01 HL096812, HL096814, HL096899, HL096902, HL096917), with previous brain MRI examinations funded by R01 HL70825 from the National Heart, Lung, and Blood Institute. This work was supported by the Framingham Heart Study (NHBLI/Nih contract HHSN268201500001) to the Boston University School of Medicine and by grants from the National Institute on Aging (R01 AG008122, R01 AG008122, R01 AG049505, U01 AG052409) and the National Institute of Neurological Disorders and Stroke (NS017950 and UH2 NS100605). Funding support for “Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium” was provided by the National Institutes of Health through the American Recovery and Reinvestment Act of 2009 (ARRA) (SRC2HL102419). Sequencing for both studies was carried out at the Baylor College of Medicine Human Genome Sequencing Center (US4 HG003273). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

variants with MAF ≤ 5%. Cross-race fixed effects meta-analyses were Bonferroni-corrected for the number of variants or genes tested.

Results

Seven genes were associated with aβ in late middle age or change over time; no associations were identified in old age. Single variants in KLKB1 (rs3733402; p = 4.33x10^-10) and F12 (rs1801020; p = 3.89x10^-6) were significantly associated with midlife aβ42 levels through cross-race meta-analysis; the KLKB1 variant replicated internally using 1,014 additional participants with exome chip. ITPRIP, PLIN2, and TSPAN18 were associated with the midlife aβ42:aβ40 ratio via the T5 test; TSPAN18 was significant via the cross-race meta-analysis, whereas ITPRIP and PLIN2 were European American-specific. NCOA1 and NT5C3B were associated with the midlife aβ42:aβ40 ratio and the fold-change in aβ42, respectively, via SKAT in African Americans. No associations replicated externally (N = 725).

Conclusion

We discovered age-dependent genetic effects, established associations between vascular-related genes (KLKB1, F12, PLIN2) and midlife plasma aβ levels, and identified a plausible Alzheimer’s Disease candidate gene (ITPRIP) influencing cell death. Plasma aβ concentrations may have dynamic biological determinants across the lifespan; plasma aβ study designs or analyses must consider age.

Introduction

Alzheimer’s disease (AD) is a major public health burden, affecting 5.1 million Americans aged 65 or older; the number of cases is expected to triple by 2050, costing the nation $1.1 trillion [1]. Although there are no effective treatments to prevent, slow, or cure AD, researchers have made considerable progress in dissecting its genetic etiology and revealing biological pathways that may contain druggable targets. The International Genomics of Alzheimer’s Project leveraged results from genome-wide association studies (GWAS) of late-onset AD (LOAD) to identify candidate therapeutic targets in the immune response, endocytosis, cholesterol transport, and protein ubiquitination pathways [2]. GWAS and sequencing studies have nominated variants (representing ~30 genes) from the whole allele frequency spectrum [3–7], yet there is a substantial proportion of LOAD heritability unexplained [6, 7]. The identification of additional therapeutic candidates is hindered by the reduced power of case-control outcomes [8].

More recent studies have employed endophenotypes of particular facets of AD pathophysiology to identify known and novel candidate genes using moderate sample sizes [9–12]. Capitalizing on the critical role that amyloid-β (aβ) plays in AD pathophysiology [12], along with functional information linking most AD-associated genes to aβ production and clearance [4], genome-wide association studies have employed cerebrospinal fluid (CSF) and plasma aβ concentrations as endophenotypes [9–12]. Plasma levels are less expensive and invasive to measure than CSF levels, are associated with brain aβ [13–16], and may capture any shared mechanisms regulating aβ processes (production, secretion, degradation, and clearance) throughout the body [12, 17]. However, their utility as AD endophenotypes is questionable. Only a handful of prospective studies and systematic meta-analyses have associated plasma
levels or the $\alpha_\beta_{42}:\alpha_\beta_{40}$ ratio with LOAD [18–23], while a paucity of associations have been documented between plasma $\alpha_\beta$ traits and AD linkage and GWAS findings [17, 24, 25]. Furthermore, plasma $\alpha_\beta$ concentrations reflect a wide array of tissue sources [26] and age-related health conditions, such as subcortical white matter lesions, cerebral microbleeds, hypertension, diabetes, infarcts, ischemic heart disease, and chronic kidney disease [14, 27, 28], a few of which are themselves associated with AD. This fosters ambiguity in the pathophysiological interpretation (AD or non-AD) of any findings.

Therefore, before exploiting plasma $\alpha_\beta$ concentrations in genetic or epidemiological studies of AD, a more comprehensive assessment of their age-dependent biological determinants is needed [27]. Deciphering the dynamic genetic architecture of plasma $\alpha_\beta$ concentrations could simultaneously implicate pathophysiological processes, pathways, and mechanisms contributing to $\alpha_\beta$ plaque accumulation in the asymptomatic preclinical and progressive phases of AD [29, 30] while suggesting novel health conditions that alter the association between plasma $\alpha_\beta$ and AD in epidemiological studies. This vital information may clarify the inconsistencies in reported associations between plasma $\alpha_\beta$ and AD across studies, as well as provide insight into the role (if any) of plasma $\alpha_\beta$ in AD pathogenesis and the utility of plasma $\alpha_\beta$ concentrations as AD biomarkers.

To systematically assess the dynamic biological determinants of plasma $\alpha_\beta$, we performed an exome-wide association study of plasma $\alpha_\beta$ concentrations measured during two different life stages. The goal was to identify genes influencing plasma $\alpha_\beta_{42}$ concentrations and $\alpha_\beta_{42}:\alpha_\beta_{40}$ ratios in late middle age, old age, or change over time. We performed single-variant and gene-based association analyses using whole exome sequence from 1,414 Atherosclerosis Risk in Communities-Neurocognitive Study (ARIC-NCS) participants. This study was well-suited for this investigation because: 1) participants had two amyloid measurements spaced an average of 18 years apart, with mean ages of 59 and 77 years for the two blood draws; 2) the sample included both European (EAs) and African Americans (AAs), allowing exploration of rare population-specific variants that may contribute to health disparities; 3) the sample was enriched for dementia and mild cognitive impairment, and hence should be enriched for variants contributing to $\alpha_\beta$ accumulation; and 4) the availability of whole exome sequence permitted the analysis of both common and rare variants, enabling the first systematic interrogation of rare variants for plasma $\alpha_\beta$ levels. This investigation expanded upon the lone published GWAS meta-analysis of plasma $\alpha_\beta$ concentrations [12] which focused on cross-sectional measurements in non-demented elderly participants of European ancestry.

**Subjects and methods**

**Subjects**

ARIC was initiated in 1987 as a population-based cohort study of 15,792 middle-aged (45–64 years) participants drawn from four US communities (Washington County, MD; Forsyth County, NC; Jackson, MS; and suburban Minneapolis, MN) [31]. Four study visits were completed by 1999, with a fifth visit (ARIC-NCS; N = 6,538) conducted in 2011–2013 [32]. Plasma $\alpha_\beta$, the phenotype for this investigation, was quantified on a subset (N = 2,588) of ARIC-NCS enriched for cognitive impairment. The plasma $\alpha_\beta$ sample included all individuals exhibiting impaired cognitive status (defined as low mini-mental status exam score or low standardized score on any of five cognitive domains accompanied by cognitive decline on longitudinally administered tests) during the fifth exam, all participants with a brain MRI from a prior ARIC exam, and an age-stratified (<80 years, ≥80 years) random sample of the remaining cognitively normal participants from each field center. This investigation focused on 1,414 AA and EA participants with whole exome sequence, covariates, and amyloids measured at the third
(1993–1995) and fifth visits. Of these participants, 152 (72 AAs and 80 EAs) had dementia and 560 (152 AAs and 408 EAs) had mild cognitive impairment by the fifth visit; 79% of these dementia and cognitive impairment cases were ascribed to AD as the primary etiology. Cognitive diagnoses were adjudicated at the fifth visit using cognitive, neurologic, and brain imaging assessments (comprehensive diagnostic details are given in [32]). Cognitive status was not available on the whole sample at the third visit. The ARIC study has been approved by the Institutional Review Board at each field center, namely Wake Forest Baptist Medical Center (Forsyth County, NC), University of Mississippi Medical Center (Jackson, MS), University of Minnesota (suburban Minneapolis, MN), and Johns Hopkins University (Washington County, MD). Participants provided written informed consent prior to each examination.

**Plasma amyloid-β ascertainment**

Amyloid quantification was performed by the Department of Molecular Pharmacology and Experimental Therapeutics at Mayo Clinic, Jacksonville, FL, from August to December 2014. The INNO-BIA assay (INNOGENETICS N.V, Ghent, Belgium) required 69 plates to measure $\alpha_42$ and $\alpha_40$ levels in both races; for each participant, the same plate was used to simultaneously measure the amyloid levels at the third and fifth visit. Beads (xMAP microspheres; conjugate 1A) bound to $\alpha_40$ and $\alpha_42$ emitted fluorescence detected by the Luminex 200 IS Total system. A five-parameter logistic regression model related the fluorescence intensities of six standards to their known amyloid concentrations. The resultant model predicted the concentrations of $\alpha_40$ and $\alpha_42$ from the measured fluorescence intensities in the samples. Intensities outside the range of the standards could not be inferred. For the single-visit analyses of $\alpha_42$, samples with intensities below the minimal detectable level were assigned the threshold concentrations (12 pg/ml). These individuals were omitted from the analysis of fold-changes in $\alpha_42$ and all analyses of the $\alpha_42:\alpha_40$ ratio since their ranks relative to those with measured values were inconclusive; 98 (36 AAs, 62 EAs), 17 (6 AAs, 11 EAs), 107 (41 AAs, 66 EAs), and 105 (41 AAs, 64 EAs) participants were omitted from the ratio at visit 3, the ratio at visit 5, the fold-change in the ratio, and the fold-change in $\alpha_42$ analyses, respectively, due to subthreshold intensities.

**Whole exome sequencing**

DNA samples were assembled into Illumina paired-end pre-capture libraries; the oligonucleotide sequences and protocol are available on the Baylor College of Medicine Human Genome Sequencing Center (HGSC) website (http://www.hgsc.bcm.edu/content/protocols-sequencing-library-construction). Two, four, or six pre-capture libraries were pooled together, hybridized to the HGSC VCRome 2.1 design [33] (42Mb, NimbleGen), and sequenced in a single lane on the Illumina HiSeq 2000 or the HiSeq 2500 platform. The HGSC Mercury pipeline (https://www.hgsc.bcm.edu/content/mercury) conducted the Illumina sequence analysis while the Consensus Assessment of Sequence and Variation program de-multiplexed the pooled samples. The Burrows-Wheeler Alignment [34] algorithm mapped reads to the Genome Reference Consortium Human Build 37 (GRCh37) sequence, producing Binary Alignment/Map (BAM) files. Aligned reads were then recalibrated using the Genome Analysis ToolKit [35], BAM sorting, duplicate read marking, and realignment near insertions or deletions (indels). The Atlas2 [36] suite called both single nucleotide variants (SNVs) and indels, generating high-quality variant call files.

The ARIC exome sequence was quality controlled as part of the Cohorts for Hearts and Aging Research in Genomic Epidemiology (CHARGE) consortium. SNVs and indels were centrally filtered for posterior probabilities<0.95, variant read counts<3, variant read ratios
<0.25 or >0.75, total read depths of the references<10-fold, mappability scores<0.8, missing rates>20%, mean coverage depths>500-fold, race-specific Hardy-Weinberg Equilibrium p-values<5x10^{-6}, and strand bias (>99% variant reads in a single strand direction). In addition, SNVs with total coverage<10-fold and indels with total coverage<30-fold were excluded. Individuals were excluded for sex-mismatch, missingness>20%, singleton counts of 0, or values>6 standard deviations from the mean for singleton counts, race-specific mean depths, TiTv ratios, or heterozygote to homozygote ratios. The mean depth of coverage was 109.6 and 82.7 for AAs and EAs, respectively.

The annotation file included quality-controlled variants observed in at least one available exome sequencing project (e.g. CHARGE, the NHLBI Exome Sequencing Project). Variants were annotated using ANNOVAR [37] and dbNSFP v2.0 according to GRCh37 and RefSeq. This multiple study SNPinfo file was used as a component of the R package seqMeta (http://cran.r-project.org/web/packages/seqMeta/index.html).

Covariates
We included age, sex, and apolipoprotein-E (APOE) as covariates in this investigation. APOE genotypes were determined using TaqMan assays and the ABI 7700 Sequence Detection System (Applied Biosystems, Foster City, CA). APOE e4 carrier status (0 = No, 1 = Yes) indicated whether an individual carried at least one copy of the e4 allele; this carrier status was strongly associated (minimum p-value of 1x10^{-8} for αβ42 at visit 5) with age- and gender-adjusted plasma αβ levels in EAs and was moderately (p-values in the range of 0.1 to 0.2) associated in AAs. Although the sample included APOE e4 homozygotes (23 (6%) and 20 (2%) in the AAs and EAs, respectively), APOE e4 carrier status tended to fit better (by Akaike Information Criterion in models of age- and gender-adjusted plasma αβ) than separate coefficients for e4 heterozygotes and homozygotes or the number of e4 alleles.

Population structure
Race-specific principal components (PCs), the first ten of which were utilized to control for population stratification in the statistical analysis, were calculated from the genotype data using Eigenstrat [38]. Variants with minor allele frequencies (MAFs)<0.05, missing rates>0.05, or Hardy-Weinberg Equilibrium p-values<1x10^{-5} were excluded before pruning the variants for linkage disequilibrium (r^2)>0.3. In total, 29,551 and 16,323 SNVs were used to construct the PCs in AAs and EAs, respectively.

Statistical methods
Single-visit amyloid measures were linearly regressed onto age, gender, and APOE e4 carrier status separately by race (field center was not statistically significant). The fold-change amyloid traits were regressed onto the age at the third visit, the time elapsed between visits, gender, and APOE e4 carrier status. Following rank-based inverse normal transformation of the residuals, we used seqMeta (version 1.5; http://cran.r-project.org/web/packages/seqMeta/) to conduct race-specific single-variant and gene-based association tests while adjusting for population structure (the first ten PCs). Linear regression models, assuming additive genetic effects, were fit on all autosomal SNVs with MAF≥0.01. T5 burden and Sequence Kernel Association (SKAT; using default “Wu” weights) gene-based tests were conducted on functional (nonsynonymous, splicing, stop gain, stop loss, or frameshift) autosomal variants with MAF ≤ 0.05. For both the single-variant and gene-based tests, we meta-analyzed the African and European American results (score statistics and genotype covariance matrices) using fixed effects models in seqMeta. We applied a Bonferroni correction for the number of unique variants (up to
113,423) or genes (up to 16,733) tested per trait (S1 Table contains the number of variants/genes tested per trait), yielding exome-wide significance thresholds of 4.41x10^{-7} (0.05/113,423) and 2.99 x10^{-6} (0.05/16,733) for the single-variant and gene-based tests, respectively. The quantile-quantile plots from the race-specific analyses and the cross-race meta-analyses exhibited minimal inflation and were well-behaved for both the single-variant (genomic inflation factors ranged from 0.99 to 1.03; S1 Fig) and gene-based tests (both overall and by minor allele count (MAC) thresholds; S2–S7 Figs).

Replication

We internally replicated significant single-variant tests using 1,014 ARIC participants with exome chip (Illumina Human Exome BeadChip v1.0) but not exome sequence data; these individuals had plasma αβ measured at both the third and fifth visits and were used to replicate both cross-sectional and change over time traits. The genotype calling and quality-control procedures have been described elsewhere [39], while the statistical methods matched those of the sequence analyses. We did not internally replicate significant gene-based tests because few functional SNVs (≤ 2) were available in the genes of interest using exome chip participants from the appropriate racial group. We externally replicated both the significant single-variant and gene-based tests using 725 participants from the Framingham Heart Study (FHS) Offspring and Third Generation (Gen 2 and Gen 3) cohorts [40, 41]; these participants had amyloids measured once and were used to replicate cross-sectional findings only. All significant cross-sectional associations from the discovery phase were with midlife plasma αβ levels (i.e. from the third visit in ARIC), thus we restricted the FHS replication sample to participants aged 50–70 years (mean age of 60±6 years, 48% female). The amyloid assessment was consistent between FHS and ARIC (the INNO-BIA assay conducted at the same lab), as was the exome sequence calling, quality control, and annotation (both members of CHARGE). The single-variant and gene-based analyses of the inverse-transformed amyloids incorporated a kinship matrix to account for FHS family structure and included significant PCs.

Results

The discovery analysis included 406 AAs and 1,008 EAs with whole exome sequence (Table 1; for summary statistics by cognitive status, see S2 and S3 Tables). Only one-third of AAs were male, whereas about half (47%) of EAs were male. Both races had similar age distributions, with the third visit corresponding to late middle age and the fifth visit corresponding to older age, and high APOE ε4 carriage rates compared to the general population (20–25%) [42]. In both AAs and EAs, mean αβ_{42} levels increased between visits and mean αβ_{42}:αβ_{40} ratios decreased between visits. Although the third visit αβ_{42} levels differed between the races, the mean fold-changes were similar. The internal (ARIC exome chip) and external (FHS exome sequence) replication samples had lower APOE ε4 carriage rates (34% for ARIC AAs, 18% for ARIC EAs, and 22% for FHS EAs) than the discovery samples (S4 Table). The FHS replication sample included only EAs and had more males (52%) and higher midlife plasma αβ levels (both αβ_{42} and αβ_{42}:αβ_{40} ratio) than all ARIC samples.

Single-variant analysis

**ARIC exome sequence analyses.** Kallikrein B, plasma (Fletcher factor) 1 [KLKB1] and coagulation factor XII (Hageman factor) [F12] contained common SNVs significantly (p≤4.41x10^{-7}) associated with αβ_{42} traits (Table 2). Nonsynonymous SNV rs3733402 in KLKB1 was associated with αβ_{42} levels at the third visit and the fold-change in αβ_{42} across visits; these associations were significant using the EAs alone but had the same directions of effects.
and moderate statistical support (p<0.09) in AAs. EAs with the A allele had lower \( a\beta_{42} \) levels at the third visit but an increased fold-change in \( a\beta_{42} \) over visits (Fig 1A); individuals with this allele may start late-midlife with lower levels but have larger increases over the next 16–20 years. A second SNV in \( KLKB1 \) (rs925453) was associated with the third visit \( a\beta_{42} \) levels but became insignificant after conditioning on rs3733402 (p-values of 0.27 and 0.43 in AAs and EAs, respectively), undermining its claim as an independent finding.

Rs1801020 in the 5’ untranslated region of \( F12 \) was significantly associated with the third visit \( a\beta_{42} \) levels through cross-race meta-analysis (Table 2). Both races provided evidence of association (p < 0.0001), with lower \( a\beta_{42} \) levels for the G allele. The lowest tertile of \( a\beta_{42} \) had the largest proportion of participants with the GG genotype in both races (Fig 1B).

Eight suggestive associations (4.41\( \times 10^{-7} < p < 8.82\times 10^{-6} \)) with \( a\beta_{42} \) were identified (S5 Table), including SNVs in \( PPP5C \) (rs917948), \( FUT9 \) (rs9499636), \( ECHS1 \) (rs2230260), \( SMAP1 \) (rs576516), \( PDZD8 \) (rs35664484), \( RRP12 \) (rs6584122), \( CACNA2D4 \) (rs202022529) and \( ADGRF5 \) (rs678312).

**Replication analyses.** We internally validated the significant \( KLKB1 \) (rs3733402) associations using 1,014 ARIC participants who had exome chip but not exome sequence data. The directions of effects on the third visit \( a\beta_{42} \) levels and the fold-change in \( a\beta_{42} \) levels were consistent between participants with exome chip and sequence (Table 2). The rs3733402 associations replicated in EAs alone, with strengthened evidence in the cross-race meta-analysis. We could not internally replicate rs1801020 in \( F12 \); it was not available on the exome chip and was

### Table 1. Descriptive statistics of sequenced ARIC participants.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>African Americans (N = 406)</th>
<th>European Americans (N = 1,008)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Female</td>
<td>67.7%</td>
<td>53.4%</td>
</tr>
<tr>
<td>% APOE ε4 carriers</td>
<td>44.1%</td>
<td>33.0%</td>
</tr>
<tr>
<td>Time between amyloid measurements (in years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>18 (1)</td>
<td>18 (1)</td>
</tr>
<tr>
<td>Range</td>
<td>16–20</td>
<td>16–20</td>
</tr>
<tr>
<td>Age (in years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visit 3 Mean (SD)</td>
<td>59 (5)</td>
<td>60 (5)</td>
</tr>
<tr>
<td>Range</td>
<td>50–71</td>
<td>50–70</td>
</tr>
<tr>
<td>Visit 5 Mean (SD)</td>
<td>77 (5)</td>
<td>77 (5)</td>
</tr>
<tr>
<td>Range</td>
<td>67–89</td>
<td>67–90</td>
</tr>
<tr>
<td>Plasma ( a\beta_{42} ) (in pg/ml)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visit 3 Mean (SD)</td>
<td>26.69 (9.30)</td>
<td>30.06 (9.64)</td>
</tr>
<tr>
<td>Interquartile Interval</td>
<td>20.10–31.80</td>
<td>23.20–36.34</td>
</tr>
<tr>
<td>Visit 5 Mean (SD)</td>
<td>33.57 (12.06)</td>
<td>38.45 (11.23)</td>
</tr>
<tr>
<td>Interquartile Interval</td>
<td>25.50–39.70</td>
<td>30.70–45.20</td>
</tr>
<tr>
<td>Fold-change Mean (SD)</td>
<td>1.30 (0.49)</td>
<td>1.31 (0.41)</td>
</tr>
<tr>
<td>Interquartile Interval</td>
<td>1.04–1.44</td>
<td>1.06–1.47</td>
</tr>
<tr>
<td>( a\beta_{42} :a\beta_{40} ) ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visit 3 Mean (SD)</td>
<td>0.20 (0.11)</td>
<td>0.19 (0.07)</td>
</tr>
<tr>
<td>Interquartile Interval</td>
<td>0.15–0.22</td>
<td>0.15–0.23</td>
</tr>
<tr>
<td>Visit 5 Mean (SD)</td>
<td>0.17 (0.07)</td>
<td>0.17 (0.06)</td>
</tr>
<tr>
<td>Interquartile Interval</td>
<td>0.13–0.20</td>
<td>0.13–0.19</td>
</tr>
<tr>
<td>Fold-change Mean (SD)</td>
<td>0.95 (0.33)</td>
<td>0.93 (0.31)</td>
</tr>
<tr>
<td>Interquartile Interval</td>
<td>0.75–1.11</td>
<td>0.73–1.08</td>
</tr>
</tbody>
</table>

NOTE: SD = Standard deviation. The amyloid measures were non-normal, thus the interquartile interval conveyed the amyloid values in the 25th to 75th percentile.

https://doi.org/10.1371/journal.pone.0180046.t001
poorly imputed in the 1000 Genomes GWAS data (AAs and EAs had imputation qualities of 0.707 and 0.606, respectively). Neither the \textit{KLKB1} or \textit{F12} variants externally replicated in FHS.

**Gene-based tests**

**ARIC exome sequence analyses.** Inositol 1,4,5-triphosphate receptor interacting protein \textit{[ITPRIP]}, perlipin-2 \textit{[PLIN2]}, tetraspanin 18 \textit{[TSPAN18]}, nuclear receptor coactivator 1 \textit{[NCOA1]}, and 5’-nucleotidase, cytosolic IIIB \textit{[NT5C3B]} were associated (p≤2.99x10\textsuperscript{-6}) with plasma αβ via gene-based tests (Table 3) and had cumulative MAFs ranging from 0.76% to 1.49% in the significant race-specific- or meta-analysis. \textit{ITPRIP}, \textit{PLIN2}, and \textit{TSPAN18} were significantly associated with the third visit αβ\textsubscript{42}:αβ\textsubscript{40} ratio via the T5 burden test; the \textit{ITPRIP} and \textit{PLIN2} associations were EA-specific with no shared variants across races, whereas \textit{TSPAN18} showed nominal association (p<0.05) in both races, had one variant (rs138778813) Table 2. Significant single-variant results.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Gene</th>
<th>SNV Rs Number (Functional Region)</th>
<th>Coded Allele</th>
<th>Chr</th>
<th>Position in base pairs (Build 37)</th>
<th>Analysis</th>
<th>Race</th>
<th>N</th>
<th>CAF</th>
<th>β</th>
<th>se(β)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>aβ\textsubscript{42} at visit 3</td>
<td>KLKB1</td>
<td>rs3733402 (Nonsynonymous)</td>
<td>A</td>
<td>4</td>
<td>187,158,034</td>
<td>ARIC Exome Sequence</td>
<td>AA</td>
<td>406</td>
<td>0.73</td>
<td>-0.18</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EA</td>
<td>1,007</td>
<td>0.51</td>
<td>-0.26</td>
<td>0.04</td>
<td>3.55E-09</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Meta</td>
<td>1,413</td>
<td>0.57</td>
<td>-0.24</td>
<td>0.04</td>
<td>4.33E-10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ARIC Exome Chip</td>
<td>AA</td>
<td>149</td>
<td>0.69</td>
<td>-0.12</td>
<td>0.14</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EA</td>
<td>865</td>
<td>0.51</td>
<td>-0.19</td>
<td>0.05</td>
<td>4.02E-05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Meta</td>
<td>1,014</td>
<td>0.54</td>
<td>-0.19</td>
<td>0.04</td>
<td>3.07E-05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Meta-ARIC</td>
<td>ALL</td>
<td>2,427</td>
<td>0.56</td>
<td>-0.22</td>
<td>0.03</td>
<td>9.90E-14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FHS</td>
<td>EA</td>
<td>725</td>
<td>0.53</td>
<td>0.04</td>
<td>0.05</td>
<td>0.41</td>
</tr>
</tbody>
</table>

| F12 | rs1801020 (UTR5) | G | 5 | 176,836,532 | ARIC Exome Sequence | AA | 406 | 0.57 | -0.28 | 0.07 | 7.71E-05 |
|     |                       |     |     |             | EA | 1,007 | 0.76 | -0.20 | 0.05 | 8.37E-05 |
|     |                       |     |     |             | Meta | 1,413 | 0.71 | -0.23 | 0.04 | 3.89E-08 |
|     |                       |     |     |             | FHS | EA | 725 | 0.75 | -0.05 | 0.06 | 0.39 |

| Fold-change in aβ\textsubscript{42} | KLKB1 | rs3733402 (Nonsynonymous) | A | 4 | 187,158,034 | ARIC Exome Sequence | AA | 364 | 0.71 | 0.14 | 0.08 | 0.09 |
|     |                       |     |     |     | EA | 941 | 0.50 | 0.27 | 0.05 | 9.28E-09 |
|     |                       |     |     |     | Meta | 1,305 | 0.56 | 0.24 | 0.04 | 5.03E-09 |
|     |                       |     |     |     | ARIC Exome Chip | AA | 132 | 0.69 | 0.19 | 0.15 | 0.20 |
|     |                       |     |     |     | EA | 818 | 0.50 | 0.16 | 0.05 | 8.88E-04 |
|     |                       |     |     |     | Meta | 950 | 0.53 | 0.16 | 0.05 | 3.80E-04 |
|     |                       |     |     |     | Meta-ARIC | ALL | 2,255 | 0.55 | 0.20 | 0.03 | 1.58E-11 |

**NOTE:** Chr = Chromosome; N = Sample size; CAF = Coded allele frequency; β = Effect of each copy of the coded allele on the trait; se(β) = Standard error of the effect of each copy of the coded allele on the trait. SNV rs1801020 was not available on the exome chip in ARIC and was poorly imputed in the 1000 Genomes GWAS data in ARIC (IMPUTE2 imputation qualities of 0.707 and 0.606 in AAs and EAs, respectively).

https://doi.org/10.1371/journal.pone.0180046.t002
shared across races, and was significant in the cross-race meta-analysis. Carrying a minor allele in \textit{ITPRIP} or \textit{TSPAN18} increased the third visit $\alpha_\beta_{42}:\alpha_\beta_{40}$ ratio in EAs whereas carrying a minor allele in \textit{PLIN2} decreased the ratio (Fig 2). These patterns were reflected in the single-variant model coefficients of the contributing variants (Fig 3 and S6–S8 Tables); all \textit{PLIN2} variants had negative coefficients in EAs and all but one of the \textit{TSPAN18} and \textit{ITPRIP} variants had positive coefficients. The lone overlapping \textit{TSPAN18} variant (rs138778813) increased the ratio in both races but showed greater evidence of association in EAs than AAs (single-variant p-values of $3.98 \times 10^{-5}$ and 0.67, respectively), possibly due to differences in the number of copies of the minor allele in the two groups (10 and 1 minor allele copies, respectively).

\textit{NCOA1} and \textit{NT5C3B} were associated with the third visit $\alpha_\beta_{42}$ levels and the fold-change in $\alpha_\beta_{42}$ levels, respectively, by SKAT (Table 3). These associations were AA-specific and had one major contributing SNV each. Rs1804645 (\textit{NCOA1}; $p = 8.54 \times 10^{-7}$; 6 minor allele copies)
and rs138262483 (NT5C3B; p = 9.72x10^-7; 4 minor allele copies) were the primary contributors (Fig 4 and S9 and S10 Tables). No NT5C3B variants overlapped races but NCOA1 had one variant (rs1804645) with opposite directions of effects and different allele frequencies in AAs and EAs (MAF of 0.008 and 0.037, respectively). Sixteen additional genes (ABL2, ATP2A1, Table 3. Significant gene-based (T5 and SKAT) results.

<table>
<thead>
<tr>
<th>Significant Test in ARIC</th>
<th>Trait</th>
<th>Gene</th>
<th>Chr</th>
<th>Analysis</th>
<th>Race</th>
<th>N</th>
<th>MAC</th>
<th># SNVs</th>
<th>T5 Test</th>
<th>SKAT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aβ42:αβ40 ratio at visit 3</td>
<td>ITPRI P</td>
<td>10</td>
<td>ARIC</td>
<td></td>
<td>AA</td>
<td>370</td>
<td>47</td>
<td>8</td>
<td>-0.11</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EA</td>
<td>945</td>
<td>20</td>
<td>5</td>
<td>1.13</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Meta</td>
<td>1,315</td>
<td>67</td>
<td>13</td>
<td>0.28</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FHS</td>
<td>EA</td>
<td>725</td>
<td>10</td>
<td>0.10</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SKAT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aβ42:αβ40 ratio at visit 3</td>
<td>NCOA1</td>
<td>2</td>
<td>ARIC</td>
<td></td>
<td>AA</td>
<td>370</td>
<td>11</td>
<td>5</td>
<td>0.90</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>EA</td>
<td>945</td>
<td>85</td>
<td>15</td>
<td>-0.11</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Meta</td>
<td>1,315</td>
<td>96</td>
<td>19</td>
<td>0.00</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FHS</td>
<td>EA</td>
<td>725</td>
<td>58</td>
<td>-0.06</td>
<td>0.13</td>
</tr>
<tr>
<td>Fold-change in aβ42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NT5C3B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fold-change in aβ42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Chr = Chromosome; N = Sample size; MAC = Number of copies of the minor alleles; # SNVs = Number of SNVs contributing to the gene-based test; \( \beta \) = Effect of the minor alleles on the trait; se(\( \beta \)) = Standard error of the effect of the minor alleles.

One of the PLIN2 SNVs (rs35568725) had a minor allele frequency of 0.059 in EAs but 0.005 in AAs, thus was included in the meta-analysis (combined minor allele frequency 0.044) for both cohorts by default; we performed the meta-analysis both including and omitting this variant in the EA group. Including this variant in both racial groups yielded a MAC of 146, a T5 meta-analysis p-value of 8.59E-05 (\( \beta = -0.33, \text{se}(\beta) = 0.08 \)), and a SKAT meta-analysis p-value of 1.05E-03. Excluding this variant in the EAs yielded a minor allele count of 35, a T5 meta-analysis p-value of 2.12E-04 (\( \beta = -0.64, \text{se}(\beta) = 0.17 \)), and a SKAT meta-analysis p-value of 3.36E-03.

Fold-change in aβ42 was not available in FHS.

https://doi.org/10.1371/journal.pone.0180046.t003

Fig 2. Effects of carrying a minor allele in TSPAN18, ITPRIP, and PLIN2 on the third visit aβ42:αβ40 ratio in EAs. The plotted values are the inverse normal transformed ratios adjusted for age, gender, and APOE ε4 carriage status.

https://doi.org/10.1371/journal.pone.0180046.g002
Fig 3. Variants contributing to the significant T5 gene-based tests. This figure depicts the single-variant model results for variants contributing to the T5 tests of ITPRIp, PLIN2, and TSPAN18 on the third visit $\alpha_4^{\beta_42}:\alpha_4^{\beta_40}$ ratio. One isoform was drawn unless the regions harboring variants differed substantially. Each contributing variant...
is labeled by name, single-variant association p-value, and direction of the minor allele effect (in parentheses). Variants with nominal (p<0.05) levels of association are in red. Overlapping variants in the two racial groups are encapsulated by a red box.

Fig 4. Variants contributing to the significant SKAT gene-based tests. This figure displays the single-variant model results for variants contributing to the SKAT tests of NCOA1 and NT5C3B on the third visit $\alpha_{4\beta}^{\beta_{4\beta}}$ ratio and fold-change in $\alpha_{4\beta_{4\beta}}$, respectively. One isoform was drawn unless the regions harboring variants differed substantially. Each contributing variant is labeled by name, single-variant association p-value, and direction of the minor allele effect. Variants with nominal (p=0.05) levels of association are in red. Overlapping variants in the two racial groups are encapsulated by a red box.

https://doi.org/10.1371/journal.pone.0180046.g004
CCDC102B, CSHL1, DSG4, FGF23, FLNB, GALNT13, LMF2, MED25, NIPAL2, NUDT7, SNAPC1, TARS2L, TMEM202, TRUB2) were suggestively (2.99 x 10^{-6} < p ≤ 5.98 x 10^{-5}) associated with amyloids through the T5 or SKAT tests (S11 and S12 Tables).

**Replication analyses.** Of the four genes associated with the third visit aβ_{42}/aβ_{40} ratio, none strictly replicated in FHS cohorts (Table 3). However, ITPRIP was nominally (p-value < 0.05) associated using SKAT instead of the T5 burden test. Rs114174365, the only ITPRIP variant with more than one minor allele copy in both ARIC and FHS EAs, had similar effect sizes (0.89 and 1.01) in both, with p-values 0.001 and 0.045, respectively (S13 Table). PLIN2, NCOA1, and TSPAN18 did not even nominally replicate. The largest contributing TSPAN18 variant (rs138778813; 10 and 21 minor allele copies in ARIC and FHS EAs) had opposite directions of effects and different allele frequencies (0.005 and 0.015, respectively) in the two cohorts, possibly due to study design issues (relateds in FHS, required survival to older age and oversampling for cognitive impairment in ARIC). We could not replicate the association between NT5C3B and the fold-change in aβ_{42} because FHS had only one amyloid assessment available.

**Temporal sensitivity of associations**

We compared the exome-wide p-values from the third and fifth visits using scatterplots but found no systemic patterns (such as the third visit yielding lower p-values than the fifth visit, etc.) for either aβ_{42} or the ratio using single-variant or gene-based tests. The ranks of the p-values at the two time points were only weakly correlated (Spearman correlation coefficients ranging from 0.14 to 0.34). Our significant and suggestive findings demonstrated temporal sensitivity; no findings were significant or suggestive at both visits (S14–S18 Tables). Only half (F12, PLIN2, NCOA1) of the significant third visit associations showed nominal evidence at the fifth visit. The effect of rs3733402 (KLKB1) on aβ_{42} levels was near zero (p-values > 0.620) for all fifth visit analyses. Rs1801020 (F12) was nominally associated with aβ_{42} in the fifth visit EA- and meta-analyses (p-values 0.014 and 0.036, respectively) but the effect estimates were attenuated (-0.126 and -0.088, respectively) compared to the third visit (-0.201 and -0.229, respectively). Similar results were observed for the significant gene-based tests.

**Phenotype-specificity of associations**

All significant associations were specific to one type of amyloid measure (aβ_{42} or the ratio; see S19 and S20 Tables) and failed to produce even suggestive evidence for the other. For the two common variants associated with aβ_{42} traits (rs3733402 in KLKB1 and rs1801020 in F12), the fold-change was the ratio measure with the strongest cross-race evidence (meta-analysis p-values of 0.005 and 1.46 x 10^{-4} and consistent direction of effects across AAs and EAs). We also note that two of the ratio-associated genes (ITPRIP and NCOA1) failed to produce nominal associations (p-value < 0.05) for any aβ_{42} trait in any analysis (AA-specific, EA-specific, or meta-analysis).

**Discussion**

Leveraging late-midlife plasma aβ concentrations and fold-changes from a moderate-sized biracial sample enabled the identification of seven genes that were obscured in cross-sectional analyses of the elderly. These findings echoed those of the published GWAS meta-analysis of plasma aβ concentrations which failed to identify significant genetic variants in non-demented elderly participants [12]. Plasma aβ concentrations change with age, at least until dementia or brain plaques appear [22, 43], and could have a dynamic genetic architecture across the life-span. Such age-dependent genetic associations have been reported for other complex traits,
including blood pressure [44, 45] and lipids [46], and can represent genes that are active at specific ages or genes that are active across the age spectrum with varying effect magnitudes [45]. Accounting for age-dependent genetic effects, whether through model-based approaches or age stratification, can enhance gene discovery efforts and our understanding of intridual variation in plasma a\(\beta\) levels [44–46]. Such analyses may suggest the optimal age range to capture particular facets of disease (e.g. AD, vascular) pathophysiology while negating the need for massive sample sizes.

Numerous phenomena could reshape the genetic landscape with aging. Gene expression and posttranslational protein modification are dynamic processes that change with age [45]. An accumulation of behavioral and environmental exposures can cause epigenetic changes, such as DNA methylation, histone modification, and microRNA expression, thereby altering gene expression (and hence the detectable genetic effects present) across the lifespan [47].

Age-related oxidative damage may mediate an accumulation of posttranslational modifications to proteins [48], altering their function and the effects of genes encoding them. More specific to our phenotypes, age-related health conditions may increase the heterogeneity of genetic effects contributing to plasma a\(\beta\) levels, thereby changing the genetic underpinnings of these traits with age; subcortical white matter lesions, cerebral microbleeds, hypertension, diabetes, infarcts, ischemic heart disease, and chronic kidney disease have been associated with plasma a\(\beta\) levels or the ratio [14, 27, 28]. Gene-age interactions could arise from differences in the a\(\beta\) processes perturbed (production, degradation, transportation across the blood brain barrier, reabsorption through the CSF, or clearance from the body [49–53]) or tissue sources (platelets, skeletal muscle, pancreas, kidney, liver, vascular wall, lung, intestine, skin, glands, and brain [26]) across the lifespan. For example, the genetic underpinnings of plasma a\(\beta\) could mimic the pattern hypothesized for early- versus late-onset AD [26]; genes involved in a\(\beta\) production may have a greater impact on plasma concentrations in the young, while genes involved in a\(\beta\) clearance may have a greater impact in the elderly.

Three gene-based associations (ITPRIP, TSPAN18, and PLIN2) had decent minor allele counts in the discovery sample and plausible connections to plasma a\(\beta\). The protein encoded by ITPRIP binds to the inositol-1,4,5-triphosphate receptor (IP\(_3\)R) and allosterically downregulates IP\(_3\)-induced calcium release from the endoplasmic reticulum [54]. Oligomeric a\(\beta\) alters IP\(_3\)-triggered calcium release [55] and, in return, IP\(_3\)-induced calcium release may influence amyloid precursor protein (APP) cleavage to a\(\beta\) and a\(\beta\) [56]. Thus, ITPRIP may interact with IP\(_3\)R to modulate both the a\(\beta\) concentration and synaptic plasticity through calcium signaling [54]. A pathway analysis of plasma a\(\beta\) GWAS results lent credibility to our ITPRIP finding [12]; eighteen of twenty-seven Ingenuity canonical pathways associated with plasma a\(\beta\) contained the receptor (IP\(_3\)R) modulated by ITPRIP. Located near a previously reported AD linkage peak [57], ITPRIP is an attractive AD candidate therapeutic target because of its role in cell death; it binds and inhibits the activity of death-associated protein kinase, a key component in cell death signaling pathways [58]. Itprip knock-out mice show increased cell death [58]; the effect on neuronal death is quite pronounced, with Itprip knock-out hippocampal neurons exhibiting 28% survival of wild-types after exposure to stimuli [58]. Reduction of IP\(_3\)R-mediated calcium signaling rescued presenilin-associated AD pathogenesis in mouse models [59], thus drugs that enhance ITPRIP activity may prevent cell death in neurodegenerative diseases and stroke [58].

TSPAN18, supported by both races in ARIC, encodes a four-transmembrane protein from a highly conserved family known to influence the spatial organization of membrane proteins through interactions with each other, signaling proteins, enzymes, transmembrane receptors, and adhesion molecules [60]. Tetraspanins also participate in cellular trafficking through endocytic and recycling organelles, lysosomes, or secreted vesicles [61, 62]. Up- or down-
regulation of tetraspanin proteins and microdomains influences αβ production, specifically through α- and γ-secretase processing of APP [63]. The tetraspanin associated with plasma αβ in this investigation, TSPAN18, has been inconsistently associated with schizophrenia in Han Chinese [64, 65], while other tetraspanins have been associated with schizophrenia, bipolar disorder, and X-linked mental retardation [66, 67].

PLIN2 encodes perilipin-2, the most prevalent lipid-droplet-associated protein in nonadipose tissue [68]. Perilipin-2 is associated with lipid droplet biogenesis, cholesterol efflux, plasma very low-density lipoprotein cholesterol levels, and intracellular and plasma triglyceride levels [68–70]. The association between αβ and PLIN2 is plausible since cholesterol concentrations and trafficking impact APP processing and αβ degradation [71, 72] and lipid droplet presence is positively correlated with αβ levels in brain neurons from AD patients [73]. The association between PLIN2 and plasma αβ may be due to atherosclerosis [70, 74, 75]; perilipin-2 aids foam cell formation [74] and is overexpressed in atherosclerotic plaques [70] which produce αβ [76]. The overexpression of PLIN2 in macrophages increases the expression of monocyte chemoattractant protein-1 [77], which itself elevates αβ accumulation in AD mice [78]. The association between PLIN2 and plasma αβ may also be due to a host of metabolic disorders, such as fatty liver disease and obesity, which are associated with perilipin-2 and αβ levels [79, 80]. We note that PLIN2 is near a well-known AD linkage peak on chromosome 9p22.1 [81].

Single-variant tests linked two key members (F12 and KLKB1) of the kallikrein-kinin system to plasma αβ. F12 and KLKB1 encode proteins that participate in complement activation, blood clotting, neutrophil aggregation, fibrinolysis (through plasminogen activation), and the bioprocessing of vasoactive peptides [82]. Most notably, contact activation of coagulation factor XII (FXII; encoded by F12) causes cleavage of plasma prekallikrein (encoded by KLKB1) to kallikrein, liberating bradykinin from high molecular weight kininogen [83]. The significant variants in F12 (rs1801020) and KLKB1 (rs3733402) have documented biological consequences. The A allele at Rs1801020, which is four bases upstream of the translation initiation codon, decreases translation efficiency and plasma FXII activity [84]. The G allele at rs3733402 substitutes serine for asparagine in an apple domain that mediates the binding of plasma prekallikrein to high molecular weight kininogen [83]. Thus, rs3733402 and rs1801020 are expected to alter bradykinin levels; they have also been associated with levels of plasma renin [85], biological surrogates of endothelin-1 and adrenomedullin [86], and B-type natriuretic peptide (rs3733402 only) [87]. The renin-angiotensin and endothelin-1 signaling pathways have been associated with plasma αβ via a pathway analysis of GWAS results [12], somewhat corroborating our common variant associations.

There are several potential mechanisms connecting the kallikrein-kinin system to αβ. Bradykinin, endothelin-1, and angiotensin-1 (a product of the renin cascade) are substrates for the αβ-degrading proteases neprilysin, endothelin-converting enzyme, and angiotensin converting enzyme, respectively [53], possibly influencing αβ degradation rates. Or, perhaps, bradykinin could participate in a feedback loop to regulate αβ levels; αβ may interact with FXII to increase bradykinin levels while bradykinin may decrease the formation of αβ through α-secretase processing of APP [88–91]. Whether the association between plasma αβ and the kallikrein-kinin system reflects any AD pathophysiology is unknown [92]. Variants in KLKB1 or F12 were not identified in the largest GWAS of AD to date [3], although the kallikrein-kinin system is overactivated in the plasma, CSF, and frontal and temporal cortices of AD cases [90, 92, 93] and is most prominently expressed in brain regions with the earliest signs of AD [94]. FXII, its binding sites, and components of its proteolytic cascades are present in αβ plaques of autopsied brains [95]. In addition, studies have shown that bradykinin can incite tau protein phosphorylation and subsequent learning and memory impairments [96], induce IP₃
accumulation and mobilization of intracellular calcium [97], and impact inducible nitric oxide synthase, resulting in cognitive impairment [94].

We surmise that KLKB1 and F12 could be associated with both vascular disease and aβ deposition, with subsequent effects on AD. Pairwise associations between kallikrein-kinin system genes (KLKB1 and F12), complex vascular traits (hypertension, myocardial infarction, and stroke), and plasma aβ (aβ42 or the ratio) concentrations have been reported [14, 27, 98–101]. In turn, midlife vascular risk factors have been linked to both cognitive decline and brain amyloid deposition in late-life [102, 103], while the latter has been associated with plasma aβ concentrations [13–16]. The association of vascular-related genes with midlife, but not late-life, plasma aβ and the association of midlife, but not late-life, vascular risk factors with late-life brain amyloid deposition [103], bolsters the assertion that genetic studies of midlife plasma aβ levels can capture contributors to late-life amyloid deposition.

Although several of the observed plasma aβ genetic associations were supported by in vitro studies and animal models, external replication is required. This is particularly true for the tenacious AA-specific findings in NT5C3B and NCOA1 which had smaller minor allele counts (S21 Table details the functions of these genes and their possible connections to aβ). Unfortunately, our replication efforts were hindered by the paucity of studies with plasma aβ and exome sequence in a large number of middle-aged (especially underrepresented minority) participants. Relying on a single sample of 725 EAs exacerbated the already difficult task of replicating single-variant and gene-based tests [6]. Few functional variants overlapped across ARIC and FHS (≤3 per gene within race, ≤1 per gene across races) and the study designs differed. ARIC included AAs and EAs who survived until the fifth examination and were oversampled for cognitive impairment. In contrast, FHS included EAs who attended the visit of interest, without enriching for future cognitive status. Thus, non-replication could indicate false positives due to the complicated ARIC sample selection strategy and attrition. As is common practice in agnostic gene identification studies, we ignored these issues in the analyses which biased the estimated effects but allowed us to find and prioritize candidate genes for further study.

Our installation had a few limitations. The blood samples from the third visit were stored for two decades before amyloid assessment whereas the fifth visit samples were only stored short term. Therefore the genetic associations observed in the third visit and the change over time may not be due to the age effects but rather genes influencing plasma aβ levels (via degradation) over long term storage. This concern is somewhat minimized by the fact that aβ levels are stable for at least one year and three freeze-thaw cycles [104] and INNO-BIA assays have been used on samples stored for one to two decades [19, 105]. Some participants, potentially containing rare aβ-associated variants with large effects, had amyloid concentrations below the minimal detection threshold. The exclusion of these participants from the ratio and fold-change analyses may have precluded the identification of additional aβ-associated genes. Our primary analysis ignored the amyloid assay plate effects, included low-frequency (0.01 < MAF ≤ 0.05) variants in the burden test, and ignored interactions between variants/gens and APOE ε4 carriage or future cognitive status. Sensitivity analyses showed that these minimally impacted the significant findings (see S22–S26 Tables for analyses incorporating batch effects, restricting to rare variants (MAF ≤ 0.01), and stratifying by APOE ε4 and cognitive impairment), although the magnitude of the single-variant effects tended to be larger among those who became cognitively impaired. Lastly, we used residuals of fold changes as a simplistic outcome in this investigation but will conduct future studies using more sophisticated longitudinal analyses.

Overall, this investigation highlighted the potential age-dependency of plasma aβ genetic associations, established connections between midlife plasma aβ levels and vascular-associated genes, and suggested novel candidate AD genes for further study. Our findings implicate
complex traits (e.g. hypertension) that are associated with both plasma αβ and AD. Therefore, we must be cognizant that differences in non-AD complex trait distributions may confound the association between plasma αβ and AD across studies. A more comprehensive understanding of the biological contributors to plasma αβ across the lifespan is critical to understand their role (if any) in AD pathogenesis and their utility as AD biomarkers.

Supporting information

S1 Fig. QQ plots for single-variant tests using ARIC participants with exome sequence. (PDF)

S2 Fig. QQ plots for T5 burden tests using ARIC participants with exome sequence. (PDF)

S3 Fig. QQ plots for SKAT using ARIC participants with exome sequence. (PDF)

S4 Fig. QQ plots for the T5 test of the third visit αβ_{42}:αβ_{40} ratio in EAs using different minor allele count thresholds (starting at 0.5% cumulative MAF). (PDF)

S5 Fig. QQ plots for the T5 test of the third visit αβ_{42}:αβ_{40} ratio from the cross-race meta-analysis using different minor allele count thresholds (starting at 0.5% cumulative MAF). (PDF)

S6 Fig. QQ plots for the SKAT test of the third visit αβ_{42}:αβ_{40} ratio in AAs using different minor allele count thresholds (starting at 0.5% cumulative MAF). (PDF)

S7 Fig. QQ plots for the SKAT test of the fold-change in αβ_{42} in AAs using different minor allele count thresholds (starting at 0.5% cumulative MAF). (PDF)

S1 Table. Number of variants or genes tested for each trait in the discovery analysis. (XLSX)

S2 Table. Mean plasma αβ levels by cognitive status in African Americans. (XLSX)

S3 Table. Mean plasma αβ levels by cognitive status in European Americans. (XLSX)

S4 Table. Descriptive statistics for the internal and external replication samples. (XLSX)

S5 Table. Suggestive single-variant tests (minor allele frequencies ≥ 0.01) using ARIC participants with exome sequence. (XLSX)

S6 Table. ITPRIP variants contributing to the T5 test of the αβ_{42}:αβ_{40} ratio at visit 3 using ARIC participants with exome sequence. (XLSX)

S7 Table. PLIN2 variants contributing to the T5 test of the αβ_{42}:αβ_{40} ratio at visit 3 using ARIC participants with exome sequence. (XLSX)
S8 Table. TSPAN18 variants contributing to the T5 test of the $\alpha\beta_{42}:\alpha\beta_{40}$ ratio at visit 3 using ARIC participants with exome sequence.
(XLSX)

S9 Table. NCOA1 variants contributing to the SKAT test of the $\alpha\beta_{42}:\alpha\beta_{40}$ ratio at visit 3 using ARIC participants with exome sequence.
(XLSX)

S10 Table. NT5C3B variants contributing to the SKAT test of the fold-change in $\alpha\beta_{42}$ using ARIC participants with exome sequence.
(XLSX)

S11 Table. Suggestive T5 gene-based associations (p-value < $5.98\times10^{-5}$ and cumulative minor allele frequency $\geq 0.5\%$ in the suggestive analysis group) using ARIC participants with exome sequence.
(XLSX)

S12 Table. Suggestive SKAT associations (P-value < $5.98\times10^{-5}$ and cumulative minor allele frequencies $\geq 0.5\%$ in the suggestive analysis group) using ARIC participants with exome sequence.
(XLSX)

S13 Table. Functional variants (> 1 minor allele copy) overlapping in ARIC and FHS participants with exome sequence; contributors to the gene-based tests.
(XLSX)

S14 Table. Temporal sensitivity of significant single-variant tests in ARIC participants with exome sequence.
(XLSX)

S15 Table. Temporal sensitivity of significant gene-based tests in ARIC participants with exome sequence.
(XLSX)

S16 Table. Temporal sensitivity of suggestive single-variant results in ARIC participants with exome sequence.
(XLSX)

S17 Table. Temporal sensitivity of suggestive T5 tests in ARIC participants with exome sequence.
(XLSX)

S18 Table. Temporal sensitivity of suggestive SKAT tests in ARIC participants with exome sequence.
(XLSX)

S19 Table. Cross-trait results for significant single-variants in ARIC participants with exome sequence.
(XLSX)

S20 Table. Cross-trait results for significant genes in ARIC participants with exome sequence.
(XLSX)

S21 Table. Functions of significant genes identified by SKAT of ARIC African Americans with exome sequence.
(XLSX)
S22 Table. Sensitivity analysis for INNO-BIA plate effects; significant single-variant tests in ARIC participants with exome sequence. (XLSX)

S23 Table. Sensitivity analysis for INNO-BIA plate effects: SKAT and T5 results using ARIC participants with exome sequence. (XLSX)

S24 Table. KLKB1 and F12 single-variant tests stratified by APOE e4 carrier status; ARIC participants with exome sequence. (XLSX)

S25 Table. KLKB1 and F12 single-variant tests stratified by visit 5 cognitive status; ARIC participants with exome sequence. (XLSX)

S26 Table. Significant T1 burden tests using ARIC participants with exome sequence. (XLSX)

Acknowledgments
We would like to extend our deepest appreciation to the participants and staffs of the Atherosclerosis Risk in Communities (ARIC) Study and the Framingham Heart Study (FHS) for their important contributions.

Author Contributions
Conceptualization: Eric Boerwinkle, Thomas H. Mosley, Jr.
Formal analysis: Jeannette Simino, Zhiying Wang, Qiong Yang.
Funding acquisition: Sudha Seshadri, Eric Boerwinkle, Thomas H. Mosley, Jr.
Investigation: Steven G. Younkin.
Project administration: Sudha Seshadri, Eric Boerwinkle, Thomas H. Mosley, Jr.
Supervision: Jan Bressler, Sudha Seshadri, Myriam Fornage, Eric Boerwinkle, Thomas H. Mosley, Jr.
Visualization: Jeannette Simino, Zhiying Wang, Jan Bressler, Myriam Fornage.
Writing – original draft: Jeannette Simino, Thomas H. Mosley, Jr.
Writing – review & editing: Jeannette Simino, Zhiying Wang, Jan Bressler, Vincent Chouraki, Sudha Seshadri, Myriam Fornage, Thomas H. Mosley, Jr.

References


41. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute’s Framingham Heart Study: design,


59. Shilling D, Muller M, Takano H, Mak DO, Abel T, Coulter DA, et al. Suppression of InsP3 receptor-mediated Ca(2+) signaling alleviates mutant presenilin-linked familial Alzheimer’s disease


