Supplementary Material

Aβ-dependent and independent genetic pathways regulating CSF tau biomarkers in Alzheimer's disease

Atul Kumar, PhD[1]; Shorena Janelidze[1]; Erik Stomrud, MD, PhD[1, 2]; Sebastian Palmqvist, MD, PhD[1, 2]; Oskar Hansson, MD, PhD[1, 2, #]; Niklas Mattsson-Carlgren, MD, PhD[1, 3, 4, #]

[1]Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden

[2]Memory Clinic, Skåne University Hospital, Malmö, Sweden

[3]Department of Neurology, Skåne University Hospital, Lund, Sweden

[4]Wallenberg Centre for Molecular Medicine, Lund University, Lund, Sweden

Shared Senior Authors

Email Addresses

Atul Kumar: atul.kumar@med.lu.se

Shorena Janelidze: shorena.janelidze@med.lu.se

Sebastian Palmqvist: sebastian.palmqvist@med.lu.se

Erik Stomrud: erik.stomrud@med.lu.se

Oskar Hansson: oskar.hansson@med.lu.se

Niklas Mattsson-Carlgren: niklas.mattsson-carlgren@med.lu.se

1 eMethods

1.1 Study participants

The study population consisted of 751 Cognitively Unimpaired (CU) elderly participants, 212 patients with Mild Cognitive Impairment (MCI) and 150 patients with Alzheimer's disease (AD) dementia from the prospective and longitudinal Swedish BioFINDER sample (clinical trial no. NCT01208675; www.biofinder.se), for which age, education, gender and biomarker data were available. Following research guidelines [1], the CU group consisted both of normal controls (N=569) and of patients who had been investigated for mild cognitive symptoms and found to have subjective cognitive decline (SCD) (N=182). The participants were recruited between September 2010 and December 2014 at three different memory clinics as previously described [2, 3]. Briefly, clinical assignment of MCI and SCD was performed after patient recruitment based on a neuropsychological battery as previously described [4]. The subjects were thoroughly assessed for their cognitive complaints by physicians with a particular interest in dementia disorders. The inclusion criteria for patients with SCD and MCI were as follows: (i) objective cognitive impairment; (ii) not fulfilling the criteria for dementia; (iii) a Mini-Mental State Examination (MMSE) score of 24–30 points; (iv) age 60–80 years; and (v) fluent in Swedish. The exclusion criteria were as follows: (i) cognitive impairment that without a doubt could be explained by a condition other than prodromal dementias; (ii) severe somatic disease; and (iii) refusing lumbar puncture or neuropsychological investigation. Cognitively normal controls were eligible for inclusion if they (i) were aged 60 years old, (ii) scored 28–30 points on the MMSE at the screening visit, (iii) did not have cognitive symptoms as

evaluated by a physician, (iv) were fluent in Swedish, (v) did not fulfill the criteria of MCI or dementia. The exclusion criteria were as follows: (i) presence of significant neurologic or psychiatric disease (e.g., stroke, Parkinson's disease, multiple sclerosis, major depression), (ii) significant systemic illness making it difficult to participate, (iii) refusing lumbar puncture and (iv) significant alcohol abuse. AD dementia patients were eligible for inclusion if they (i) fulfilled the criteria of dementia due to Alzheimer's disease. The exclusion criteria were as follows: (i) Significant unstable systemic illness or organ failures, such as terminal cancer, that makes it challenging to participate in the study, (ii) Current significant alcohol or substance misuse. The Regional Ethics Committee in Lund, Sweden, approved the study. All subjects gave their written informed consent.

Genotyping and preparation of genetic data

Genotyping was conducted using the Illumina platform GSA-MDA v2. Before imputation, subject-level quality control (QC) included removing sexual incompatibility between chip-inferred sex and self-reported sex, low call rates (1% cut-off), and extreme heterozygosity. Relatedness among the samples was eliminated by removing one participant from each pair of close relatives (first or second degree) identified as $\hat{\pi} \geq 0.0625$. Using PLINK2 [5], multi-dimensional scaling was done to create principal components in genetic analyses to account for ancestry.

Standard QC steps were performed for SNP-level filter to ensure conformity with the reference panel used for imputation (strand continuity, names of the alleles, position, and assignments for Ref / Alt). To sum up, for

imputation, 685494 high-quality variants (autosomal, non-monomorphic, bi-allelic variants with Hardy–Weinberg Equilibrium (HWE) $P > 5 \times 10^{-8}$ and with a call rate of > 99%) were used.

Imputation was carried out using the Sanger Imputation Server (https://imputation.sanger.ac.uk/) with SHAPEIT for phasing [6], Positional Burrows-Wheeler Transform (PWBT) [7] for imputation, and the entire Haplotype Reference Consortium (release 1.1) reference panel [8].

Multi-allelic variants and SNPs with a data imputation score < 0.2 have been excluded as part of post-imputation QC, and genotype calls with a posterior likelihood < 0.9 have been set to fail (i.e., hard-called). SNPs with a genotyping rate >0.9 were retained. SNPs with Minor Allele Frequency (MAF) ≥ 5% were taken for the analysis (rather than, e.g., MAF >1 %, due to the relatively small sample size of the cohort). Further information on the imputation and QC process is detailed in https://rpubs.com/maffleur/452627.

Preparation of ADNI genetic data

SNP genotyping data were available for n = 1674 subjects across all ADNI phases. Three separate platforms were used for genotyping: Human610-Quad (ADNI 1), HumanOmniExpress (ADNI GO/2) (Illumina; [9]). The genome coordinates for ADNI 1 were referenced to NCBI build 36 (UCSC hg 18). These coordinates were liftover to NCBI build 37 (UCSC hg 19) using the Lift Genome Annotations tool of the UCSC genome browser [10]. The standard QC, imputation, and post imputation QC were applied separately on each data set (ADNI 1, ADNI 2, and ADNI GO). Further, the three data were merged, and

SNPweights [11] were used to infer genetic origins from genotyped SNPs using a reference panel consisting of samples from HapMap 3 (The International HapMap 3 Consortium, 2010) from northern and western Europe, Yoruba Africans, and Eastern and Southern Asian population. Subjects with 80 percent or more of expected central European ancestors were retained. The first ten PCA components were computed for these subjects in PLINK2 [5] and were included in the data to account for the population structure.

Polygenic Score Calculation

Using an $r^2 < 0.1$ threshold over 1000 kb sliding windows (index threshold and clumped SNPs of $p < 1$), linkage-disequilibrium (LD) clumping was performed using PLINK's clump function. LD clumping ensures that large blocks of correlated SNP sets do not overload PRS/PGS calculation. The *APOE* gene is the most significant AD risk factor, with strong linkage disequilibrium (LD) levels in the locus region. Therefore, SNPs falling within the *APOE* gene region (chr19:44400000-46500000; GRCh37 / hg19 assembly) were omitted from the dataset. The *APOE* ε2/ε3/ε4 status SNPs (rs7412 and rs429358) genotypes were then reintroduced into the dataset to ensure that the genetic risk of *APOE* was captured. The PRS / PGS was determined for each subject by summing up the adequate number of alleles (0, 1, 2) of the SNPs weighted by the natural logarithm of their respective ORs (Odds Ratio). The default formula for PRS calculation in PLINK is:

$$PRS_j = \frac{\sum_i^N S_i * G_{ij}}{P * M_j}$$

Where the effect size of SNP *i* is S_i; the number of effect alleles observed in sample j is G_{ij}; the ploidy of the sample is P (is generally 2 for humans); the total number of SNPs included in the PRS is N, and the number of non-missing SNPs observed in sample j is M_j. If the sample has a missing genotype for SNP i, then the population minor allele frequency multiplied by the ploidy ($MAF_i * P$) is used instead of G_{ij}.

Publicly accessible summary statistics from reported GWAS studies (not overlapping with our dataset) of AD [e12] was used to define PRS for AD. We iterated over a variety of values (0.05 to 5×10^{-8}) to evaluate the appropriate p-value threshold $p = 0.05$ [PRS 1], $p = 5 \times 10^{-3}$ [PRS 2], $p = 5 \times 10^{-4}$ [PRS 3], $p = 5 \times 10^{-5}$ [PRS 4], $p = 5 \times 10^{-6}$ [PRS 5], $p = 5 \times 10^{-7}$ [PRS 6] and the GWAS-level significance thresholds of $p = 5 \times 10^{-8}$ [PRS 7] creating models named PRS 1-7.

Bioinformatics Analysis

We tested for gene enrichment within different PRSs using gene ontology-biological process (GO-BP) terms [e13] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway terms [e14]. We also tested for gene-disease association using the DisGeNet database [e15]. Finally, the AllEnricher application was used to perform gene-function annotation [e16].

eResults

Gene enrichment in non-*APOE*-PRS2 and Aβ-dependent and independent subsets

The PRS2 (1742 SNPs) had the strongest association with the tau biomarkers in this study based on the BioFINDER cohort. We mapped the 1742 SNPs of PRS2 to 1548 different genes (eTable 36). We performed gene-function annotation to elucidate the biological mechanisms that these SNPs may modulate. We also performed the same biological enrichment analysis for PRS2-Incl-1683 (1683 SNPs) and PRS2-R-Incl-19 (19 SNPs).

PRS2 genes were enriched for 147 GO-BP terms at Bonferroni corrected p-value < 0.05. The top hit BP terms were "Dendrite morphogenesis" (10 genes) and "RNA Processing" (5 genes) (efigure 4 and eTable 37). The KEGG pathway enrichment for the genes of PRS2 returned 22 enriched terms. The top term in this category was "Phosphatidylinositol signaling system" (18 genes) (efigure 5 and eTable 38). The functional annotation for the gene-disease association found 102 enriched terms for the genes of PRS2. Most of the genes of PRS2 were related to neuropsychiatric disorders (including "Schizophrenia", "autism spectrum disorders") along with "Late-Onset Alzheimer's Disease" (efigure 6 and eTable 39).

The "PRS2-Incl-1683" (the Aβ-independent subset of PRS2) genes were enriched for 147 GO-BP terms. The top hit GO-BP terms were similar to that of PRS2. Still, four BP terms were lost in this restricted PRS set compared to full PRS2: "Amyloid-Beta Clearance", "Regulation of Nodal Signalling Pathway", "Intermembrane Sterol Transfer" and "Wnt Signalling Pathway". For the restricted Aβ-independent tau specific PRS, we also observed gene enrichment for additional terms that were not present in the full PRS 2: "Negative Regulation Of Neuron Projection Development", "Phosphatidylinositol-Mediated Signalling" and "Vascular Endothelial Growth

Factor Receptor Signalling Pathway" (eTable 40). The KEGG pathway enrichment returned 27 enriched terms. All the enriched terms for this restricted PRS were the same as of the full PRS2 with additional enrichment for five terms: "Insulin secretion", "Inflammatory mediator regulation of TRP channels", "Long-term depression", "Adherens junction" and "Cholinergic synapse" (eTable 41). For the functional annotation of gene-disease association, we did not observe any notable difference between the enriched disease terms (eTable 42).

For the Aβ-dependent PRS (PRS2-R-Incl-19), only one term ("Lipid Transport") was enriched for the GO-BP category (eTable 43). We could not find any significant enrichment for KEGG pathway terms for this PRS. For the functional annotation of gene-disease association, eight terms were enriched, with the top terms being "Familial Alzheimer Disease" and "Amyloid Plaque" (eFigure 7 and eTable 44).

eDiscussion

We found the most robust results for non-*APOE*-PRS2 in BioFINDER and conducted detailed analyses of gene enrichment in PRS2 and the two restricted PRSs (PRS2-Incl-1683 [Aβ-independent PRS] and PRS2-R-Incl-19 [Aβ-dependent PRS]). Our gene enrichment analysis for PRS 2 genes revealed "Dendrite morphogenesis" as the top GO biological process term. The complexity and diversity of dendrites are notably well recognized, and accumulating evidence suggests that the alterations in the dendrite structure are associated with many neurodegenerative diseases [e17, e18]. Thus, changes in expression or function of the proteins involved in "Dendrite

morphogenesis" might induce pathological changes in neural circuits that predispose to, or cause, neurological diseases [e17, e18]. Furthermore, the GO-BP term "Amyloid-Beta Clearance", which was enriched in the overall PRS2, was not enriched in the restricted Aβ-independent PRS (PRS2-Incl-1683), which further confirms that this restricted PRS might be tau specific and Aβ-independent. The genes enriched for "Amyloid-Beta Clearance" in the full PRS2 are *MME, MARCO, MSR1,* and *INSR*. Different studies have reported their implication in AD, as a mutation in either of these four genes restricts their Aβ degradation activity [e19, e20, e21, e22].

The pathway enrichment analysis for the genes of PRS2 and the restricted PRS2 (PRS2-Incl-1683) returned the term "Phosphatidylinositol signaling system" as the top hit. Phosphatidylinositol metabolism is required for various intracellular signaling pathways, and Aβ oligomer affects this via activating SHIP2 via the FcγRIIb receptor [e23]. Altered levels of phosphatidylinositol and the actin cytoskeletal network by amyloid plaques and extracellular Tau seeds obstruct microglial signaling pathways in Alzheimer's disease [e24]. Studies have linked this pathway term to the pathogenesis of AD, potentially by effects on the neurotransmitter signal transduction [e25].

We did not observe any notable difference between the enriched disease terms for PRS2 and restricted PRS2 (PRS2-Incl-1683) for the functional annotation of gene-disease association. Most disease terms were related to neurodegenerative disease. But for the Aβ-dependent PRS set (PRS2-R-Incl-19), two terms ("Amyloid Plaque" and "Amyloidosis") were explicitly enriched.

The enrichment of these two terms specifically for this PRS supports our finding that the genes involved contribute towards abnormal Aβ formation. These results confirm and strengthen the use of this PRS to study Aβ-dependent genetic effects (beyond the *APOE* region) on tau metabolism.

eReference

[e1] Jack Jr CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J, Liu E. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia. 2018 Apr;14(4):535-62.

[e2] Mattsson N, Insel PS, Palmqvist S, Stomrud E, Van Westen D, Minthon L, Zetterberg H, Blennow K, Hansson O. Increased amyloidogenic APP processing in APOE ε4-negative individuals with cerebral β-amyloidosis. Nature communications. 2016 Mar 7;7(1):1-7.

[e3] Janelidze S, Stomrud E, Palmqvist S, Zetterberg H, Van Westen D, Jeromin A, Song L, Hanlon D, Hehir CA, Baker D, Blennow K. Plasma β-amyloid in Alzheimer's disease and vascular disease. Scientific reports. 2016 May 31;6(1):1-1.

[e4] Petrazzuoli F, Vestberg S, Midlöv P, Thulesius H, Stomrud E, Palmqvist S. Brief Cognitive Tests Used in Primary Care Cannot Accurately Differentiate Mild Cognitive Impairment from Subjective Cognitive Decline. Journal of Alzheimer's Disease. 2020 May 11(Preprint):1-1.

[e5] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015 Dec 1;4(1):s13742-015.

[e6] Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nature methods. 2012 Feb;9(2):179-81.

[e7] Durbin R. Efficient haplotype matching and storage using the positional Burrows–Wheeler transform (PBWT). Bioinformatics. 2014 May 1;30(9):1266-72.

[e8] McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y. A reference panel of 64,976 haplotypes for genotype imputation. Nature genetics. 2016 Oct;48(10):1279-83.

[e9] Saykin AJ, Shen L, Foroud TM, Potkin SG, Swaminathan S, Kim S, Risacher SL, Nho K, Huentelman MJ, Craig DW, Thompson PM. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. Alzheimer's & dementia. 2010 May 1;6(3):265-73.

[e10] Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome research. 2002 Jun 1;12(6):996-1006.

[e11] Chen CY, Pollack S, Hunter DJ, Hirschhorn JN, Kraft P, Price AL. Improved ancestry inference using weights from external reference panels. Bioinformatics. 2013 Jun 1;29(11):1399-406.

[e12] Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, Van Der Lee SJ, Amlie-Wolf A, Bellenguez C. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nature genetics. 2019 Mar;51(3):414-30.

[e13] Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, Basu S, Chisholm RL, Dodson RJ, Hartline E, Fey P. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Research. 2021 Jan 8;49(D1): D325-34.

[e14] Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Research. 2021 Jan 8;49(D1): D545-51.

[e15] Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic acids research. 2020 Jan 8;48(D1): D845-55.

[e16] Zhang D, Hu Q, Liu X, Zou K, Sarkodie EK, Liu X, Gao F. AllEnricher: a comprehensive gene set function enrichment tool for both model and non-model species. BMC bioinformatics. 2020 Dec;21(1):1-6.

[e17] Kweon JH, Kim S, Lee SB. The cellular basis of dendrite pathology in neurodegenerative diseases. BMB reports. 2017;50(1):5.

[e18] Arikkath J. Molecular mechanisms of dendrite morphogenesis. Frontiers in cellular neuroscience. 2012 Dec 28;6:61.

[e19] Madani R, Poirier R, Wolfer DP, Welzl H, Groscurth P, Lipp HP, Lu B, Mouedden ME, Mercken M, Nitsch RM, Mohajeri MH. Lack of neprilysin suffices to generate murine amyloid-like deposits in the brain and behavioral deficit in vivo. Journal of neuroscience research. 2006 Dec;84(8):1871-8.

[e20] Yu Y, Richard DY. Microglial Aβ receptors in Alzheimer's disease. Cellular and molecular neurobiology. 2015 Jan 1;35(1):71-83.

[e21] Li B, Chen M, Aguzzi A, Zhu C. The role of macrophage scavenger receptor 1 (Msr1) in prion pathogenesis. Journal of Molecular Medicine. 2021 Mar 23:1-1.

[e22] da Costa IB, de Labio RW, Rasmussen LT, Viani GA, Chen E, Villares J, Turecki G, Smith MD, Payao SL. Change in INSR, APBA2 and IDE gene expressions in brains of Alzheimer's disease patients. Current Alzheimer Research. 2017 Jul 1;14(7):760-5.

[e23] Strosznajder JB. Phosphoinositides kinases in brain aging and amyloid beta neurotoxicity. Journal of Neurochemistry. 2003 Jun;85:11-.

[e24] Kam TI, Park H, Gwon Y, Song S, Kim SH, Moon SW, Jo DG, Jung YK. FcγRIIb-SHIP2 axis links Aβ to tau pathology by disrupting phosphoinositide metabolism in Alzheimer's disease model. Elife. 2016 Nov 11;5:e18691.

[e25] Lo Vasco VR. The Phosphoinositide signal transduction pathway in the pathogenesis of Alzheimer's disease. Current Alzheimer Research. 2018 Apr 1;15(4):355-62.