

## Post-GWAS multiomic functional investigation of the *TNIP1* locus in Alzheimer's disease implicates mediation through *GPX3*

Daniel J. Panyard<sup>1,2\*</sup>, Lianne M. Reus<sup>3,4,5</sup>, Muhammad Ali<sup>6,7,8</sup>, Jihua Liu<sup>9,10</sup>, Yuetiva K. Deming<sup>2,11,12</sup>, Qiongshi Lu<sup>9,10</sup>, Gwendlyn Kollmorgen<sup>13</sup>, Ivonne Suridjan<sup>14</sup>, Norbert Wild<sup>13</sup>, Pieter J. Visser<sup>3,4,15,16</sup>, Lars Bertram<sup>17,18</sup>, Henrik Zetterberg<sup>19,20,21,22,23</sup>, Kaj Blennow<sup>19,20</sup>, Johan Gobom<sup>19,20</sup>, Dan Western<sup>6,7,8</sup>, Yun Ju Sung<sup>6,7,8</sup>, Cynthia M. Carlsson<sup>11,12,24,25</sup>, Sterling C. Johnson<sup>11,12,24,25</sup>, Sanjay Asthana<sup>11,12,25</sup>, Carlos Cruchaga<sup>6,7,8</sup>, Betty M. Tijms<sup>3,4</sup>, Corinne D. Engelman<sup>2</sup>, Michael P. Snyder<sup>1\*</sup>

1 Department of Genetics, Stanford University School of Medicine, Stanford University, Stanford, CA 94305, United States of America

2 Department of Population Health Sciences, University of Wisconsin-Madison, 610 Walnut Street, 707 WARF Building, Madison, WI 53726, United States of America

3 Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, The Netherlands

4 Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands

5 Center for Neurobehavioral Genetics, University of California, Los Angeles, Los Angeles, California, United States of America

6 Department of Psychiatry, Washington University School of Medicine, St Louis, MO 63110, United States of America

7 NeuroGenomics and Informatics Center, Washington University School of Medicine, St Louis, MO 63110, United States of America

8 Hope Center for Neurological Disorders, Washington University School of Medicine, St Louis, MO 63110, United States of America

9 Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, WARF Room 201, 610 Walnut Street, Madison, WI 53726, United States of America

10 Department of Statistics, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706, United States of America

11 Wisconsin Alzheimer's Disease Research Center, University of Wisconsin-Madison, 600 Highland Avenue, J5/1 Mezzanine, Madison, WI 53792, United States of America

12 Department of Medicine, University of Wisconsin-Madison, 1685 Highland Avenue, 5158 Medical Foundation Centennial Building, Madison, WI 53705, United States of America

13 Roche Diagnostics GmbH, Penzberg, Germany

14 Roche Diagnostics International Ltd, Rotkreuz, Switzerland

15 Department of Psychiatry, Maastricht University, Maastricht, the Netherlands

16 Department of Neurobiology, Care Sciences and Society, Division of Neurogeriatrics, Karolinska Institutet, Stockholm, Sweden

17 Lübeck Interdisciplinary Platform for Genome Analytics, Institutes of Neurogenetics and Cardiogenetics, University of Lübeck, Lübeck, Germany

18 Department of Psychology, University of Oslo, Oslo, Norway

19 Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden

20 Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden

21 Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK

22 UK Dementia Research Institute at UCL, London, UK

23 Hong Kong Center for Neurodegenerative Diseases, Hong Kong, China

24 Wisconsin Alzheimer's Institute, University of Wisconsin-Madison, 610 Walnut Street, 9<sup>th</sup> Floor, Madison, WI 53726, United States of America

25 William S. Middleton Memorial Veterans Hospital, 2500 Overlook Terrace, Madison, WI 53705, United States of America

\*Corresponding authors

Michael Snyder, [mpsnyder@stanford.edu](mailto:mpsnyder@stanford.edu)

Daniel Panyard, [dpanyard@stanford.edu](mailto:dpanyard@stanford.edu)

## Abstract

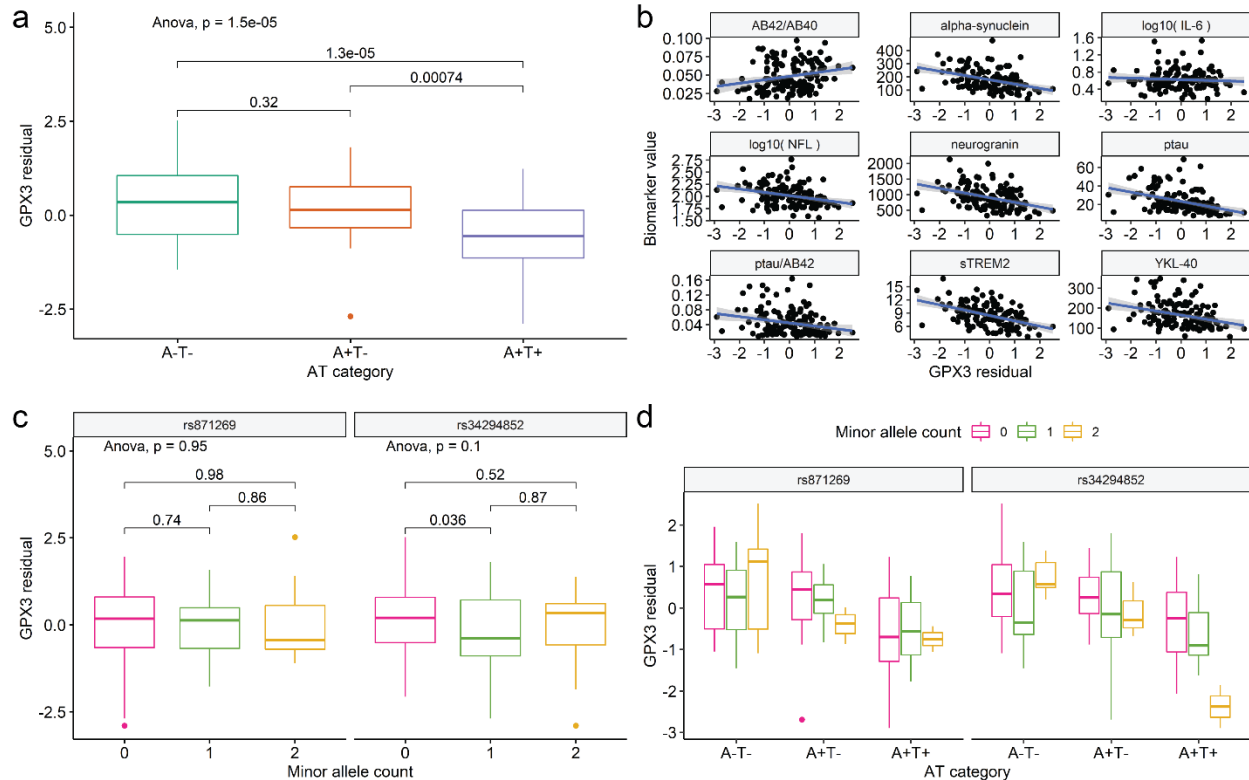
The recently reported *TNIP1/GPX3* locus from AD GWAS studies was investigated. Using proteomics and other functional omics data, we identified evidence for a functional mechanism linking variants in this locus to decreased CSF GPX3 levels as AD progresses, suggesting a new potential target for intervention.

## Main

Genomic studies in Alzheimer's disease (AD) have identified dozens of genetic associations<sup>1-4</sup>, but connecting variants with downstream drug targets is challenging. Individual functional validation to determine the consequences of mutations in implicated genes may be costly and time-consuming. In recent years, AD research studies have collected a variety of multiomics data on a large scale, including transcriptomics, proteomics, and metabolomics. These molecular data sets can provide key intermediate information linking genes to AD risk.

Several recent large-scale GWAS reported and discussed the rs871269 variant within an intron of the *TNIP1* gene (chromosome [chr] 5, base pair 151052827; all coordinates given in GRCh38 unless otherwise noted) as a protective variant for AD, with an odds ratio for AD case status of 0.96 (0.95-0.97) for the T allele, according to the Bellenguez et al. 2022 meta-analysis<sup>5,6</sup>. Wightman et al. additionally mention the rs34294852 variant (chr5:151053886), which is in low linkage disequilibrium (LD) with rs871269 in the East Asian ( $R^2 = 0.29$ ) and American ( $R^2 = 0.22$ ) populations but in lower LD ( $R^2 = 0.14$ ) among Europeans<sup>7</sup>. In terms of the functional connection of *TNIP1* with AD, Wightman et al. noted that there was little previous discussion regarding *TNIP1* and AD, but that of the three genes in the region of the variant (*TNIP1*, *GPX3*, and *SLC36A1*), *TNIP1* seemed to have the strongest connection through its regulation by Bcl3 in mice microglia and the corresponding BCL3 gene, which had been associated with CSF amyloid-beta levels in humans. Bellenguez et al. noted the connection of *TNIP1* with TNF-alpha, which was connected with the linear ubiquitin chain assembly complex (LUBAC), which itself is implicated in AD-relevant biology, including inflammation, microglia, and autophagy. Here, we use proteomics data from three different AD cohorts and annotation databases to investigate an alternative functional relationship of this locus to AD through the *GPX3* gene.

Using newly generated mass spectrometry cerebrospinal fluid (CSF) proteomics data from the University of Wisconsin (mean age 66.1; 59.9% female; detailed description previously reported<sup>8</sup>), we searched for evidence whether either *TNIP1* or *GPX3* levels were associated with AD. Specifically, we conducted ANOVA analyses that examined the relationship of these two proteins in CSF with CSF amyloid- and CSF tau-defined categories of AD (total n = 137; 56 A-T-, 39 A+T-, and 42 A+T+; see Methods). We used these amyloid (A) and tau (T) categories since they are the major biomarkers and central to research-based definitions of the disease, where amyloid tends to change first followed by tau in the current model of AD (note that A-T+ is generally excluded as presumably non-AD pathology)<sup>9,10</sup>. *TNIP1* was not been detected in the CSF in the University of Wisconsin cohorts, but *GPX3* (after correction for age and sex; see Methods) showed a statistically significant difference across the AD continuum from A-T- to A+T- to A+T+ (ANOVA  $P = 1.5 \times 10^{-5}$ ) with a significant decrease between the A-T- and A+T+ categories (t-test  $P = 1.3 \times 10^{-5}$ ) (Figure 1a). *GPX3* was also significantly associated with 8 of the 9 tested individual markers of neurodegeneration and neuroinflammation (the  $A\beta_{42}/A\beta_{40}$  ratio, phosphorylated tau [ptau], the ptau/ $A\beta_{42}$  ratio, alpha-synuclein, neurofilament light chain [NFL], neurogranin, YKL-40, and soluble TREM2 [sTREM2]), in each case negatively correlated with the biomarker except for  $A\beta_{42}/A\beta_{40}$  (Figure 1b).

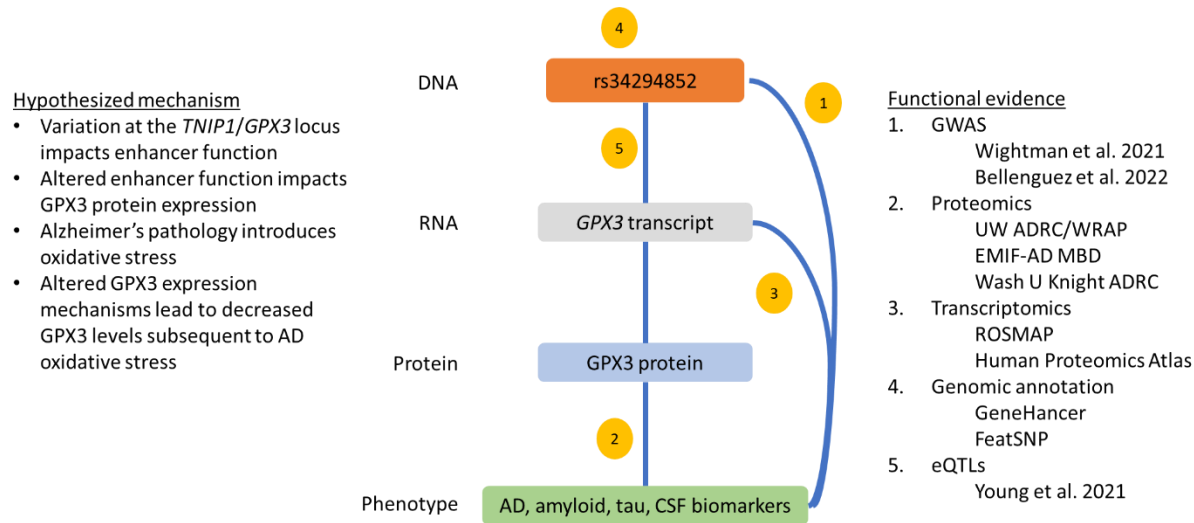


**Figure 1: Associations of CSF GPX3 with AD-related measures in the Wisconsin ADRC and WRAP cohorts. a)** The distribution of CSF GPX3 levels (after regressing out the effects of age and sex) significantly decreased across amyloid and tau (AT) positivity categories ( $n = 137$ ). **b)** CSF GPX3 levels were significantly associated with all CSF biomarkers of neurodegeneration and neuroinflammation except for IL-6 ( $n = 137$ ). In each case, GPX3 levels decreased as biomarker values indicated a worsening clinical profile. **c)** Across the cohorts ( $n = 137$ ), no difference in GPX3 levels were observed by genotype of either AD-related variant. **d)** CSF GPX3 levels by both AT and genotype were summarized for both relevant SNPs at the *TNIP1/GPX3* locus. Among participants who were A+T+ ( $n = 42$ ), CSF GPX3 levels were significantly decreased for homozygous carriers of the rs34294852 allele ( $n = 5$  among A-T-;  $n = 3$  among A+T-;  $n = 2$  among A+T+).

We then examined the genotype-mediated effects on the GPX3 trajectory. Visually and analytically (via ANOVA and pairwise t-tests), there were no differences in GPX3 levels by just the minor allele count of either SNP alone (ANOVA P values were 0.95 and 0.10 respectively across genotypes for rs871269 and rs34294852; Figure 1c), indicating that no differences in GPX3 level were present by genotype across the population. However, when we viewed these protein levels and genotypes along with AD biomarker categories, a dose-response relationship emerged for the rs34294852 genotype. When GPX3 levels were analyzed as the outcome in a multiple linear regression with both AT category and SNP minor allele count (numeric coding) as predictors, a significant decrease in GPX3 levels per copy of the minor allele was observed ( $P = 0.041$  for rs34294852; Figure 1d), suggesting that the *TNIP1/GPX3* locus-AD diagnosis relationship reported by GWAS might be mediated through changes in GPX3 level that might only become apparent when tau is abnormal in AD and particularly with homozygosity in the minor allele of rs34294852. In other words, changes in GPX3 levels might only be most evident for A+T+ individuals who are homozygous recessive for rs34294852.

Based on these results, we developed a hypothesis concerning the role of GPX3 in AD. GPX3 is a secreted glutathione peroxidase that works to protect the body from oxidative damage by reducing hydroperoxides<sup>11</sup>. AD has long been linked to oxidative stress, as have aging processes in general<sup>12,13</sup>.

The observed decline in GPX3 levels in individuals with AD might reflect insufficient or deteriorating capacity to manage oxidative stress by those with AD. This effect might be exacerbated among carriers of the minor allele of rs34294852, which presumably decreases expression of GPX3 by affecting an enhancer region interacting with the *GPX3* gene. Given the rarity of the minor allele, the subtlety of the GWAS signal, and the presence of the SNP-protein association only in the context of AD, this mechanism might be hard to detect and perhaps only then under certain disease conditions (Figure 2).



**Figure 2: Proposed functional mechanism of the *TNIP1* locus in AD.** Our hypothesis for a functional mechanism connecting the variant rs34294852 within *TNIP1* to AD outcomes is overlaid onto a map of the major omics data and the central dogma of biology. Major lines of post-GWAS functional evidence supporting this hypothesis are summarized in the right-hand list.

Were this hypothesis true, we would expect to see concordant evidence from AD transcriptomics experiments. First, we sought to understand which central nervous system cell types express *GPX3* in general. Across the cell types in the Human Protein Atlas, *GPX3* is most highly expressed in the proximal tubular cells of the kidney (20125.9 nTPM) and the Müller glia cells of the eye (5019.8 nTPM). Among brain tissues, the overall expression was lower, but microglia had the highest expression (3.7 nTPM) among the non-neuronal cell types, followed by astrocytes (1.0 nTPM) and oligodendrocytes (0.4 nTPM) (Supplementary Figure 1)<sup>14</sup>. Within an AD cohort, where *GPX3* might be more relevant given the heterogeneity of the microglia transcriptome<sup>15</sup>, RNA-seq data from the prefrontal cortex from the Religious Orders Study Memory and Aging Project (ROSMAP) showed that *GPX3* transcript levels decreased in the prefrontal cortex from controls to AD diagnosis ( $P = 7 \times 10^{-6}$ ; Supplementary Figure 2), but not in other brain regions<sup>16</sup>.

We also searched for evidence from functional genomics and transcriptomics experiments supporting a relationship between rs34294852 and *GPX3* transcription. In terms of genome functional annotation, rs34294852 is located within the 6<sup>th</sup> intron of *TNIP1* and downstream of the enhancer region GH05J151051 from GeneHancer<sup>17,18</sup>. Variant rs34294852 is predicted to alter the binding of the transcription factor MZF1 according to FeatSNP<sup>19</sup>, which is a database that aggregates brain-specific epigenetic data to examine the effects of genetic variants. In terms of expression quantitative trait loci (eQTL) data, variant rs34294852 is an eQTL for *TNIP1* and *GPX3* in blood, and it is also an eQTL for *TNIP1* in monocytes and neutrophils and for *GPX3* for neutrophils<sup>20-22</sup>. Recognizing the relatively higher expression of *GPX3* in microglia compared to other non-neuronal cell types and the implications of

microglia in AD, we investigated recent eQTL data for microglia and found that rs34294852 was an eQTL for GPX3 ( $P = 0.038$ )<sup>23</sup>.

With the evidence from the original AD GWAS and our follow-up CSF proteomics and biomarkers experiments, these additional functional and molecular data supported our central hypothesis: genetic variation at the *TNIP1/GPX3* locus impacts expression of *GPX3* through impaired enhancer activity that in turn exacerbates the observed decrease in GPX3 levels as amyloid and tau accumulate in AD progression. In cohort studies, this pathway might be difficult to detect for several reasons: 1) the original GWAS effect was small; 2) the pathway might only be relevant to a subset of cell types (e.g., microglia); and 3) and the effect might be harder to observe for homozygous dominant or heterozygous rs34294852 genotypes. Moreover, the effect allele for rs34294852 (C) has a minor allele frequency ranging from 0.16 for East Asians on the low end to 0.25 for African/African Americans on the high end<sup>7</sup>, which makes homozygous recessive individuals fairly uncommon, as was the case in the data sets analyzed here. Indeed, our attempts to replicate the proteomics signal with other population cohorts were unsuccessful. Using data from the European Medical Information Framework-Alzheimer's Disease Multimodal Biomarker Discovery (EMIF-AD MBD) study ( $n = 242$  participants), we performed an ANOVA analysis to see if GPX3 levels were different by AT category. No such difference was observed ( $P = 0.96$ ), even when stratified by rs34294852 genotypes (Supplementary Figure 3) or controlling for age, sex, or study site. Similarly, no GPX3-AD association was observed using the Knight ADRC's discovery and replication CSF proteomics data sets at Washington University in St. Louis ( $n = 1,168$  and 597 participants, respectively). The differential abundance of GPX3 levels was not significantly different between A+T+ and A-T- individuals ( $P = 0.90$  and  $P = 0.51$ , respectively) (Supplementary Figure 4). Differences in populations, proteomics technologies (mass spectrometry vs. aptamer-based), and amyloid and tau phenotyping might have affected the replication in addition to the other challenges noted above.

Nevertheless, the observed multiomic evidence connecting variation at this locus to AD through GPX3 expression, combined with our understanding of oxidative stress in AD and GPX proteins' role in combating such stress, provide a compelling hypothesis for an alternative functional mechanism at this locus that may occur in concert with any effects through TNIP1. Several functional experiments would be a reasonable next step in exploring this hypothesis: 1) using a gene-editing experiment to validate the impact of genetic variation at rs34294852 on *GPX3* expression; 2) using proteomics analysis on a microglia cell line to validate GPX3 expression; and 3) using a cell or organismal model of AD to assess the longitudinal expression of GPX proteins in connection with oxidative stress burden and its relationship to changes in tau level. More broadly, the post-GWAS analyses here demonstrate the utility of multiomic cohort data in the investigation of GWAS loci and their mechanisms of action, which can lead to new insights into Alzheimer's disease with the ultimate goal of identifying new therapeutic targets.



## Methods

### *University of Wisconsin ADRC and WRAP cohorts*

The discovery proteomics data set came from the University of Wisconsin Alzheimer's Disease Research Center (ADRC)<sup>24</sup> and Wisconsin Registry for Alzheimer's Prevention (WRAP)<sup>25</sup> cohorts, which have been described in detail previously<sup>8</sup>. This study was performed as part of the GeneRations Of WRAP (GROW) study, which was approved by the University of Wisconsin Health Sciences Institutional Review Board. Participants in the ADRC and WRAP studies provided written informed consent. Briefly, the data used here came from CSF samples taken from lumbar punctures (LPs) performed on middle- and older-aged participants in the ADRC and WRAP cohorts. All CSF samples were assayed between March 2019 and January 2020 at the Clinical Neurochemistry Laboratory at the University of Gothenburg. CSF biomarkers were assayed using the NeuroToolKit (NTK) (Roche Diagnostics International Ltd, Rotkreuz, Switzerland), a panel of automated Elecsys® and robust prototype immunoassays designed to generate reliable biomarker data that can be compared across cohorts. Measurements with the following immunoassays were performed on a cobas e 601 analyzer (Roche Diagnostics International Ltd, Rotkreuz, Switzerland): Elecsys  $\beta$ -amyloid (1–42) CSF (A $\beta$ 42), Elecsys Phospho-Tau (181P) CSF (ptau), and Elecsys Total-Tau CSF,  $\beta$ -amyloid (1–40) CSF (A $\beta$ 40), and interleukin-6 (IL-6). The remaining NTK panel was assayed on a cobas e 411 analyzer (Roche Diagnostics International Ltd, Rotkreuz, Switzerland), including markers of synaptic damage and neuronal degeneration (neurogranin, neurofilament light protein [NFL], and alpha-synuclein) and markers of glial activation (chitinase-3-like protein 1 [YKL-40] and soluble triggering receptor expressed on myeloid cells 2 [sTREM2]). Participants were categorized as amyloid positive (A+) and tau positive (T+) based on the CSF amyloid (A $\beta$ 42/A $\beta$ 40 threshold = 0.046) and tau (ptau threshold = 24.8 pg/mL) measurements, as determined through previous work<sup>26</sup>. The fourth possible category of amyloid negative and tau positive (A-T+) was not included in this study as these samples were considered to represent non-AD pathological change<sup>9</sup>. A total of nine established CSF biomarkers for AD were analyzed in this study: the A $\beta$ 42/A $\beta$ 40 ratio, ptau (pg/mL), the ptau/A $\beta$ 42 ratio, NFL (pg/mL), alpha-synuclein (pg/mL), neurogranin (pg/mL), YKL-40 (ng/mL), sTREM2 (ng/mL), and IL-6 (pg/mL). NFL and IL-6 were log<sub>10</sub>-transformed to better normalize them due to right skew.

Genotyping was performed on these participants using DNA from whole blood samples. ADRC samples genotyped by the Alzheimer's Disease Genetics Consortium (ADGC) at the National Alzheimer's Coordinating Center (NACC) using the Illumina HumanOmniExpress-12v1\_A, Infinium HumanOmniExpressExome-8 v1-2a, or Infinium Global Screening Array v1-0 (GSAMD-24v1-0\_20011747\_A1) BeadChip assay, while WRAP samples were genotyped with the Illumina Multi-Ethnic Genotyping Array at the University of Wisconsin Biotechnology Center<sup>27</sup>. Strict quality control (QC) steps and imputation to the Michigan Imputation Server<sup>28</sup> and the Haplotype Reference Consortium (HRC) reference panel<sup>29</sup> were performed. All individuals used in this study were of European ancestry. Throughout this paper, genomic coordinates are referred to by the GRCh38 genome build.

Proteomics data were generated on the CSF samples using an in-house single-shot nano-liquid chromatography-tandem mass spectrometry (nLC-MS/MS) method<sup>8,30</sup> and the raw data quantified with MaxQuant<sup>31,32</sup> using fast LFQ and a full human proteome with isoforms downloaded from UniProt



(downloaded June 14, 2017). Protein abundance data were extracted from the LFQ intensity from the “proteinGroups.txt” output file. Strict QC steps were applied to the data, yielding a final set of 915 quantified proteins for the 137 participants. All time-variable data (age, proteomics, CSF biomarkers) were quantified on the same LP sample per participant; there was no sample time discrepancy between any of the data points for an individual.

The effects of age at sample LP and sex were first regressed out in a linear regression model ( $GPX3 \sim \text{age} + \text{sex}$ ), leaving the residuals for GPX3, which was the main variable of analysis in this data set. First, an ANOVA analysis was performed across the three AT-defined categories to look for a difference in GPX3 level, with pairwise t-tests also performed for each pair of AT categories and the results visualized with box plots. Second, GPX3 residual-biomarker associations were estimated for each biomarker separately using a linear regression model. A Bonferroni correction for the number of biomarkers tested ( $P = 0.05 / 9 = 0.0056$ ) was used as the threshold for significance. Each biomarker was plotted against the GPX3 residuals, with a fitted regression line showing the relationship.

For genotype-stratified analysis, the box plots of GPX3 by AT category were repeated but with additional stratification by the count of minor alleles (T allele) for rs34294852. A linear regression was also performed with GPX3 residuals regressed on an indicator variable for A+T-, an indicator for A+T+, and the minor allele count for the variant (coded 0/1/2).

### *GPX3 eQTL analysis*

Microglia eQTL summary statistics were accessed through the European Genome-Phenome Archive (EGAD00001005736) and the Wellcome Sanger Institute Data Access Committee. All sequence datasets were aligned to human genome assembly GRCh38. Simple linear regression was used to map eQTLs with twenty-five principal components (PCs). The eQTL associations for *GPX3* were extracted for the analysis in this paper. Study designs and method details of microglia eQTL mapping have been described elsewhere<sup>23</sup>.

### *EMIF-AD MBD cohort*

The European Medical Information Framework for Alzheimer’s Disease Multimodal Biomarker Discovery (EMIF-AD MBD) study is a European multicenter study that aims to identify novel biomarkers for diagnosis and prognosis in the prodementia stages of AD<sup>33</sup>. EMIF-AD MBD combines existing clinical data and samples of 1,218 individuals with normal cognition (controls), mild cognitive impairment (MCI), or mild dementia from prospective cohort studies. The study collected baseline clinical data, MRI scans, plasma, DNA, CSF samples, and follow-up diagnosis.

DNA was extracted locally at the collection site ( $n = 805$ ) or from whole blood ( $n = 148$ ) using QIAamp DNA Blood Mini Kit (QIAGEN GmbH, Hilden, Germany) at the University of Lübeck. Genetic samples were genotyped with the Illumina Global Screening array (GSA) custom content (Illumina, Incl)

at the Institute of Clinical and Medical Biology (UKSH, Campus-Kiel). Quality control and imputation (HRC panel<sup>29</sup>) have been described in depth elsewhere<sup>34,35</sup>.

CSF was collected with lumbar puncture, as described previously<sup>33</sup>. CSF A $\beta$ <sub>1-42</sub> and total tau were measured locally with INNOTEST ELISA or INNOBIA AlzBio3 INNOBIA AlzBio3 (Fujirebio, Ghent, Belgium). Cut-offs for A $\beta$ <sub>1-42</sub> and total tau were cohort-specific in EMIF-AD. Because EMIF-AD centers had used different approaches to determine A $\beta$ <sub>1-42</sub> cut-offs, potentially leading to center-specific bias, we determined for each cohort specific cut-offs using unbiased Gaussian mixture modelling<sup>36</sup>. CSF GPX3 proteins levels were measured using tandem mass tag (TMT) mass spectrometry and have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier 10.6019/PXD019910<sup>37</sup>. GPX3 CSF protein levels were Z-score normalized with controls with normal A $\beta$ <sub>1-42</sub> and normal total tau as reference group. A full description on the quality check and normalization procedures have been described elsewhere<sup>34,38,39</sup>.

A total of n = 242 subjects had genetic, AT category (CSF A $\beta$ <sub>1-42</sub> and total tau) and CSF proteomic data (i.e., GPX3) available (n = 242). Since normal CSF amyloid and abnormal CSF total tau (A-T+) probably reflects neurodegenerative disorders other than AD, we excluded subjects with A-T+ (n = 7) from further analysis, resulting in a total of n = 235 subjects (age in years mean (SD): 66.9 (8.1); female n (%): 127 (54%); AT category: 58 A-T-, 59 A+T-, 118 A+T+; diagnostic status: 105 controls, 64 MCI, 66 AD).

GWAS analysis on GPX3 CSF levels was performed using PLINK software (v1.9) with covariate adjustment for population structure (PC1-3), age, and sex<sup>40</sup>. ANOVA analysis was performed to examine CSF GPX3 differences across AT categories, including separate analyses that additionally included age and sex or age, sex, and study site as covariates.

Written informed consent was obtained from all participants or surrogates, and the procedures for this study were approved by the institutional review boards of all participating institutions, including the following (see Bos et al 2018 for full listing<sup>33</sup>): Aristotle University of Thessaloniki Medical School Ethics Committee; Ethics Committee of the Medical Faculty Mannheim, University of Heidelberg; Ethic and Clinical Research Committee Donostia; Ethics committee Inserm and Aix Marseille University; The Healthcare Ethics Committee of the Hospital Clínic; Central Clinical Research and Clinical Trials Unit (UICEC Sant Pau); INSERM Ethical Committee; Ethic Committee of the IRCCS San Giovanni di Dio FBF; Comitato Etico IRCCS Pascale - Napoli; Ethics Committee at Karolinska Institutet; Ethische commissie onderzoek UZ/KU Leuven; Research Ethics Committee Lausanne University Hospital; Medical ethical committee Maastricht University Medical Center; Committee on Health Research Ethics, Region of Denmark; Ethics committee of Mediterranean University; University of Lille Ethics committee; Ethical Committee at the Medical Faculty, Leipzig University; Ethical Committee at the Medical Faculty, University Hospital Essen; Ethics committee University of Antwerp; Ethical Committee of University of Genoa; Ethics Committee, University of Gothenburg; Human ethics Committee of the University of Perugia; and the Medical ethics committee VU Medical Center.

*Washington University in St. Louis cohorts*

A total of 7,584 analytes were measured across 3,065 samples using SomaLogic's Somascan platform. Participants were enrolled in the Memory and Aging Project (MAP) at the Knight Alzheimer's Disease Research Center (Knight ADRC; n = 948), Alzheimer's Disease Neuroimaging Initiative (ADNI; n = 758), the Dominantly Inherited Alzheimer Network (DIAN; n = 495), Pau (n = 232), and Ruiz (n = 632) studies. All participants provided informed consent to allow their data and biospecimens to be included. The study was approved by an Institutional Review Board at Washington University School of Medicine in St. Louis.

The expression level of 7,584 proteins was measured using a multiplexed, single-stranded DNA aptamer assay developed by SomaLogic. The protein levels were reported as relative units of intensity (RFU or Relative Fluorescence Unit). Initial data normalization was performed by SomaLogic using hybridization controls for intra-plate and median signal to account for inter-plate variances<sup>41,42</sup> as well as normalization against an external reference to control for biological variances. A stringent quality check<sup>43</sup> (QC) was performed in the normalized data matrix. At the end of the QC, 316 analytes and 83 subjects were removed, resulting in a final matrix with 7,268 analytes and 2,982 subjects.

CSF samples were collected in the morning after an overnight fast, processed, and stored at -80 °C. CSF data processing for each cohort are described in detail in the respective studies<sup>44,45</sup>. For CSF samples, case-control status was determined by the Clinical Dementia Rating (CDR) at the time of lumbar puncture. For this study, the CSF A $\beta$ 42 (A) and pTau181 (T) levels were used to perform the AT classification and biomarker positivity and negativity were used to perform the differential abundance analysis. These log-normalized CSF biomarker levels (AT) were used for dichotomizing each participant into biomarker positive (case) and negative (control), as previously described<sup>44</sup>.

A multiple linear regression model was used to identify differentially abundant analytes between discovery (MAP and Ruiz; 497 AT+ and 671 AT- participants) and replication cohorts (ADNI and Pau; 360 AT+ and 237 AT- participants) where age, gender, plate ID, and the first two surrogate variables (SVs) were used as the covariates.

## Figures

Figure 1: Associations of CSF GPX3 with AD-related measures in the Wisconsin ADRC and WRAP cohorts

Figure 2: Proposed functional mechanism of the *TNIP1* locus in AD

## Supplementary Figures

Supplementary Figure 1: *GPX3* transcript expression levels in different brain cell types

Supplementary Figure 2: *GPX3* transcript expression levels across different brain regions

Supplementary Figure 3: CSF GPX3 levels by AT category and rs34294852 genotype in the EMIF-AD MBD cohort

Supplementary Figure 4: CSF GPX3 levels by AT category in the MAP/Ruiz and ADNI/Pau cohorts

## Acknowledgments

We would like to thank the participants and staff of the WRAP, Wisconsin ADRC, EMIF-AD MBD, and Knight ADRC. Without their efforts this research would not be possible.

## Funding

This research is supported by National Institutes of Health (NIH) grants R01AG27161 (Wisconsin Registry for Alzheimer Prevention: Biomarkers of Preclinical AD), R01AG054047 (Genomic and Metabolomic Data Integration in a Longitudinal Cohort at Risk for Alzheimer's Disease), P41GM108538 (National Center for Quantitative Biology of Complex Systems), R01AG037639 (White Matter Degeneration: Biomarkers in Preclinical Alzheimer's Disease), R01AG021155 (The Longitudinal Course of Imaging Biomarkers in People at Risk of AD), R21AG067092 (Identifying Metabolomic Risk Factors in Plasma and CSF for Alzheimer's Disease), and P50AG033514 and P30AG062715 (Wisconsin Alzheimer's Disease Research Center Grant), the Clinical and Translational Science Award (CTSA) program through the NIH National Center for Advancing Translational Sciences (NCATS) grant UL1TR000427, and the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. Computational resources were supported by a core grant to the Center for Demography and Ecology at the University of Wisconsin-Madison (P2CHD047873). We also acknowledge use of the facilities of the Center for Demography of Health and Aging at the University of Wisconsin-Madison, funded by NIA Center grant P30AG017266. Author LMR was funded by the Memorabel fellowship "Identifying biological and clinical relevance of (epi)genetic risk factors in sporadic FTD" (ZonMW project number: 10510022110012). Author YKD was supported by a training grant from the National Institute on Aging (T32AG000213). Author PJV was supported by grants from the European Commission, IMI (AMYPAD: 115952; RADAR-AD: 806999; and EPND: 101034344) and the ZonMW (Redefining Alzheimer's disease, #733050824736, and NCDC, #73305095005). Author HZ is a Wallenberg Scholar supported by grants from the Swedish Research Council (#2018-02532), the European Research Council (#101053962), Swedish State Support for Clinical

Research (#ALFGBG-720931), the Alzheimer Drug Discovery Foundation (ADDF), USA (#201809-2016862), the AD Strategic Fund and the Alzheimer's Association (#ADSF-21-831376-C, #ADSF-21-831381-C and #ADSF-21-831377-C), the Olav Thon Foundation, the Erling-Persson Family Foundation, Stiftelsen för Gamla Tjänarinnor, Hjärnfonden, Sweden (#FO2019-0228), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860197 (MIRIADE), and the UK Dementia Research Institute at UCL (#UKDRI-1003). Author KB is supported by the Swedish Research Council (#2017-00915), the Swedish Alzheimer Foundation (#AF-930351, #AF-939721, and #AF-968270), Hjärnfonden, Sweden (#FO2017-0243 and #ALZ2022-0006), the Swedish state under the agreement between the Swedish government and the County Councils, the ALF-agreement (#ALFGBG-715986 and #ALFGBG-965240), and the Alzheimer's Association 2021 Zenith Award (ZEN-21-848495). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Author JG is supported by Alzheimerfonden (AF-930934) and the Foundation of Gamla Tjänarinnor. Author CC receives support from the National Institutes of Health (R01AG044546, R01AG064877, RF1AG053303, R01AG058501, U01AG058922, R01AG064614, 1RF1AG074007), and the Chuck Zuckerberg Initiative (CZI). The recruitment and clinical characterization of research participants at Washington University were supported by NIH P30AG066444, and P01AG003991. This work was supported by access to equipment made possible by the Hope Center for Neurological Disorders, the NeuroGenomics and Informatics Center (NGI: <https://neurogenomics.wustl.edu/>) and the Departments of Neurology and Psychiatry at Washington University School of Medicine. Authors LMR, PJV, and BMT are supported by the ZonMW Memorabel grant programma (#733050824), and author PJV is additionally supported by the Innovative Medicines Initiative Joint Undertaking under the EMIF grant agreement (#115372). ELECSYS, COBAS and COBAS E are trademarks of Roche. The Roche NeuroToolKit robust prototype assays are for investigational purposes only and are not approved for clinical use. We thank the University of Wisconsin Madison Biotechnology Center Gene Expression Center for providing Illumina Infinium genotyping services. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

### **Author contributions**

Author DJP conceived the idea for the study; conducted the analyses of the various Wisconsin data sets; prepared the figures and tables; and led the writing of the manuscript. Author LMR conducted the analyses in the EMIF-AD MBD cohort. Author MPS helped conceive the idea for the study and interpret the results. Author MA conducted the analyses in the Knight ADRC cohort. Authors JL and QL provided the eQTL results in microglia. Author YKD contributed to the interpretation of the results. Authors HZ, KB, GK, IS, and NW contributed to the generation of the University of Wisconsin NTK CSF biomarker data set. Authors CDE, CMC, SCJ, SA, HZ, and KB contributed resources or funding. Authors CMC, SCJ, and SA assisted with the Wisconsin Alzheimer's disease cohort studies and the data and analyses conducted there for this work. Authors DW, YS, and CC assisted with the Knight ADRC cohorts and the data and analyses conducted there for this work. Authors PJV, LB, HZ, KB, JG, and BMT assisted with the EMIF-AD MBD cohorts and the data and analyses conducted there for this work. All authors contributed to and critically reviewed the manuscript.

## Competing interests

Author CC receives research support from Biogen, Eisai, Alector, GSK and Parabon; these funders of the study had no role in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. Author CC is a member of the advisory board of Vivid Genomics, Halia Therapeutics and ADx Healthcare. Author HZ has served at scientific advisory boards and/or as a consultant for Abbvie, Alector, Annexon, Apellis, Artery Therapeutics, AZTherapies, CogRx, Denali, Eisai, Nervgen, Novo Nordisk, Pinteon Therapeutics, Red Abbey Labs, Passage Bio, Roche, Samumed, Siemens Healthineers, Triplet Therapeutics, and Wave, has given lectures in symposia sponsored by Cellectricon, Fujirebio, Alzecure, Biogen, and Roche, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program. Author KB has served as a consultant, at advisory boards, or at data monitoring committees for Abcam, Axon, BioArctic, Biogen, Julius Clinical, Lilly, MagQu, Novartis, Roche Diagnostics, and Siemens Healthineers, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program. Author GK is a full-time employee of Roche Diagnostics GmbH. Author IS is a full-time employee and shareholder of Roche Diagnostics International Ltd. Author NW is a full-time employee of Roche Diagnostics GmbH. Author SCJ serves as a consultant to Roche Diagnostics and receives research funding from Cerveau Technologies. Author MPS is a cofounder and scientific advisor of Personalis, SensOmics, Qbio, January AI, Fodsel, Filtricine, Protos, RTHM, Iollo, Marble Therapeutics and Mirvie. He is a scientific advisor of Genapsys, Jupiter, Neuvivo, Swaza, and Mitrix.

Other authors have no competing interests to declare.

## Data and materials availability

The data sets analyzed from the Wisconsin ADRC and WRAP studies may be requested at <https://www.adrc.wisc.edu/apply-resources>. Microglia eQTL summary statistics may be requested through the European Genome-Phenome Archive (EGAD00001005736) and the Wellcome Sanger Institute Data Access Committee. The EMIF-AD proteomics data may be requested from the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier 10.6019/PXD019910. The Knight ADRC proteomic data is available at NIAGADS: NG00102 collection and can be interactively explored at [http://ngi.pub:3838/ONTIME\\_Proteomics/](http://ngi.pub:3838/ONTIME_Proteomics/).

## References

1. Naj, A. C. *et al.* Common variants in MS4A4/MS4A6E, CD2uAP, CD33, and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* **43**, 436–441 (2011).
2. Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452–1458 (2013).
3. Jansen, I. E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* **51**, 404–413 (2019).
4. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nat Genet* **51**, 414 (2019).
5. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* **54**, 412–436 (2022).
6. Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* **53**, 1276–1282 (2021).
7. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
8. Panyard, D. J. *et al.* Large-scale proteome analysis of CSF implicates altered glucose metabolism in Alzheimer's disease. *medRxiv* (2021) doi:10.1101/2021.09.02.21262642.
9. Jack, C. R. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* **14**, 535–562 (2018).
10. Jack, C. R. *et al.* Update on hypothetical model of Alzheimer's disease biomarkers. *Lancet Neurol* **12**, 207–216 (2013).
11. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–D745 (2016).



12. Tönnies, E. & Trushina, E. Oxidative Stress, Synaptic Dysfunction, and Alzheimer's Disease. *J Alzheimers Dis* **57**, 1105–1121 (2017).
13. Behl, C. Alzheimer's disease and oxidative stress: implications for novel therapeutic approaches. *Progress in Neurobiology* **57**, 301–323 (1999).
14. Karlsson, M. *et al.* A single-cell type transcriptomics map of human tissues. *Science Advances* **7**, eabh2169.
15. Masuda, T., Sankowski, R., Staszewski, O. & Prinz, M. Microglia Heterogeneity in the Single-Cell Era. *Cell Rep* **30**, 1271–1281 (2020).
16. Morabito, S., Miyoshi, E., Michael, N. & Swarup, V. Integrative genomics approach identifies conserved transcriptomic networks in Alzheimer's disease. *Human Molecular Genetics* **29**, 2899–2919 (2020).
17. Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)* **2017**, (2017).
18. Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* **54**, 1.30.1-1.30.33 (2016).
19. Ma, C., Madden, P., Gontarz, P., Wang, T. & Zhang, B. FeatSNP: An Interactive Database for Brain-Specific Epigenetic Annotation of Human SNPs. *Frontiers in Genetics* **10**, (2019).
20. Kwong, A. *et al.* FIVEx: an interactive eQTL browser across public datasets. *Bioinformatics* btab614 (2021) doi:10.1093/bioinformatics/btab614.
21. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. *Cell* **167**, 1398-1414.e24 (2016).
22. Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. 447367 Preprint at <https://doi.org/10.1101/447367> (2018).

23. Young, A. M. H. *et al.* A map of transcriptional heterogeneity and regulatory variation in human microglia. *Nat Genet* **53**, 861–868 (2021).
24. Melah, K. E. *et al.* CSF markers of Alzheimer’s pathology and microglial activation are associated with altered white matter microstructure in asymptomatic adults at risk for Alzheimer’s disease. *J Alzheimers Dis* **50**, 873–886 (2016).
25. Johnson, S. C. *et al.* The Wisconsin Registry for Alzheimer’s Prevention: A review of findings and current directions. *Alzheimers Dement (Amst)* **10**, 130–142 (2017).
26. Van Hulle, C. A. *et al.* An examination of a novel multipanel of CSF biomarkers in the Alzheimer’s disease clinical and pathological continuum. *Alzheimer’s & Dementia* **17**, 431–445 (2021).
27. Darst, B. F., Lu, Q., Johnson, S. C. & Engelman, C. D. Integrated analysis of genomics, longitudinal metabolomics, and Alzheimer’s risk factors among 1,111 cohort participants. *Genet Epidemiol* **43**, 657–674 (2019).
28. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284–1287 (2016).
29. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279–1283 (2016).
30. McKetney, J. *et al.* Pilot proteomic analysis of cerebrospinal fluid in Alzheimer’s disease. *Proteomics Clin Appl* e2000072 (2021) doi:10.1002/prca.202000072.
31. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372 (2008).
32. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nature Protocols* **11**, 2301–2319 (2016).
33. Bos, I. *et al.* The EMIF-AD Multimodal Biomarker Discovery study: design, methods and cohort characteristics. *Alzheimer’s Research & Therapy* **10**, 64 (2018).

34. Gobom, J. CSF proteomic study of Alzheimer's Disease. (2020).
35. Hong, S. *et al.* Genome-wide association study of Alzheimer's disease CSF biomarkers in the EMIF-AD Multimodal Biomarker Discovery dataset. *Transl Psychiatry* **10**, 1–12 (2020).
36. Tijms, B. M. *et al.* Unbiased Approach to Counteract Upward Drift in Cerebrospinal Fluid Amyloid- $\beta$  1–42 Analysis Results. *Clinical Chemistry* **64**, 576–585 (2018).
37. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research* **47**, D442–D450 (2019).
38. Tijms, B. M. *et al.* Pathophysiological subtypes of Alzheimer's disease based on cerebrospinal fluid proteomics. *Brain* **143**, 3776–3792 (2020).
39. Batth, T. S., Francavilla, C. & Olsen, J. V. Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *J. Proteome Res.* **13**, 6176–6186 (2014).
40. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742-015-0047–8 (2015).
41. Gold, L. *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLoS One* **5**, e15004 (2010).
42. Candia, J. *et al.* Assessment of Variability in the SOMAscan Assay. *Sci Rep* **7**, 14248 (2017).
43. Yang, C. *et al.* Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. *Nat Neurosci* 1–11 (2021) doi:10.1038/s41593-021-00886-6.
44. Deming, Y. *et al.* Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol* **133**, 839–856 (2017).
45. Maxwell, T. J. *et al.* Genome-wide association study for variants that modulate relationships between cerebrospinal fluid amyloid-beta 42, tau, and p-tau levels. *Alzheimer's Research & Therapy* **10**, 86 (2018).