RESEARCH

Open Access

Multivariate GWAS of Alzheimer's disease CSF biomarker profiles implies *GRIN2D* in synaptic functioning

Alexander Neumann[1,2,3†], Olena Ohlei[4†], Fahri Küçükali[1,2], Isabelle J. Bos[5], Jigyasha Timsina[6,7], Stephanie Vos[8], Dmitry Prokopenko[9], Betty M. Tijms[10], Ulf Andreasson[11,12], Kaj Blennow[11,12], Rik Vandenberghe[13,14], Philip Scheltens[10], Charlotte E. Teunissen[15], Sebastiaan Engelborghs[2,16], Giovanni B. Frisoni[17], Oliver Blin[18], Jill C. Richardson[19], Régis Bordet[20], Alberto Lleó[21,22], Daniel Alcolea[21,22], Julius Popp[23,24], Thomas W. Marsh[6,7,25], Priyanka Gorijala[6,7], Christopher Clark[23], Gwendoline Peyratout[24], Pablo Martinez-Lage[26], Mikel Tainta[26,27], Richard J. B. Dobson[28,29,30,31,32], Cristina Legido-Quigley[33,34], Christine Van Broeckhoven[2,35], Rudolph E. Tanzi[9], Mara ten Kate[36,37], Christina M. Lill[4,38,39], Frederik Barkhof[40,41], Carlos Cruchaga[6,7,42], Simon Lovestone[43,44], Johannes Streffer[2,45,46], Henrik Zetterberg[11,12,47,48,49], Pieter Jelle Visser[2,8,36], Kristel Sleegers[1,2], Lars Bertram[4,50*] and EMIF-AD & ADNI study group

Abstract

Background Genome-wide association studies (GWAS) of Alzheimer's disease (AD) have identified several risk loci, but many remain unknown. Cerebrospinal fluid (CSF) biomarkers may aid in gene discovery and we previously demonstrated that six CSF biomarkers (β-amyloid, total/phosphorylated tau, NfL, YKL-40, and neurogranin) cluster into five principal components (PC), each representing statistically independent biological processes. Here, we aimed to (1) identify common genetic variants associated with these CSF profiles, (2) assess the role of associated variants in AD pathophysiology, and (3) explore potential sex differences.

Methods We performed GWAS for each of the five biomarker PCs in two multi-center studies (EMIF-AD and ADNI). In total, 973 participants ($n = 205$ controls, $n = 546$ mild cognitive impairment, $n = 222$ AD) were analyzed for 7,433,949 common SNPs and 19,511 protein-coding genes. Structural equation models tested whether biomarker PCs mediate genetic risk effects on AD, and stratified and interaction models probed for sex-specific effects.

Results Five loci showed genome-wide significant association with CSF profiles, two were novel (rs145791381 [inflammation] and *GRIN2D* [synaptic functioning]) and three were previously described (*APOE*, *TMEM106B*, and *CHI3L1*). Follow-up analyses of the two novel signals in independent datasets only supported the *GRIN2D* locus, which contains several functionally interesting candidate genes. Mediation tests indicated that variants in *APOE* are associated with AD status via processes related to amyloid and tau pathology, while markers in *TMEM106B* and *CHI3L1* are associated with AD only via neuronal injury/inflammation. Additionally, seven loci showed sex-specific associations with AD biomarkers.

---

†Alexander Neumann and Olena Ohlei are equal first authors.

*Correspondence:
Lars Bertram
lars.bertram@uni-luebeck.de
Full list of author information is available at the end of the article

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 2 of 23

Conclusions These results suggest that pathway and sex-specific analyses can improve our understanding of AD genetics and may contribute to precision medicine.

Keywords Alzheimer's disease, Dementia, Biomarkers, Cerebrospinal fluid (CSF), Genome-wide association study (GWAS), Multivariate analysis, Principal component analysis, Mediation, Structural equation modeling

Background

Alzheimer'sdisease (AD) is a genetically complex disorder to which various pathophysiological processes are thought to contribute. Amyloid and tau pathology are the most well-known, but other processes, such as inflammation and cholesterol metabolism, among many others, play important roles in disease development as well [1]. Different risk factors may affect AD development by different mechanisms; therefore, patients may develop AD due to different combinations of causes and pathways. Accurately identifying and distinguishing which molecular mechanisms play the lead role on an individual basis is therefore crucial for etiological research, but also for clinical diagnosis, prognosis, and future therapeutic approaches.

Cerebrospinal fluid (CSF) biomarkers can provide insights into disease mechanisms, often before symptoms fully develop [2]. We have previously demonstrated the utility of linearly combining different AD CSF biomarkers into five statistically independent components, which likely represent different disease processes and which may be more informative than analyzing each CSF trait separately [3]. Specifically, we had applied principal component analysis (PCA) to data for six CSF biomarkers collected in two independent cohorts: the European Medical Information Framework for Alzheimer's Disease Multimodal Biomarker Discovery (EMIF-AD MBD) study [4] and the Alzheimer's Disease Neuroimaging Initiative (ADNI) [5].

A very similar structure representing five principal components (PC) was found in both cohorts and can be summarized as follows: [3] the first PC loaded strongly on tau and phosphorylated tau (pTau), and moderately on neurogranin (Ng) and YKL-40. Tau is a marker of neurodegeneration, with pTau being a component of neurofibrillary tangles [2, 6], Ng is a marker of synaptic functioning [7], while YKL-40 is associated with neuronal inflammation and astroglial reaction [2, 8–10]. Thus, this PC likely represents tau pathology and associated degenerative processes, such as deficits in synaptic functioning and elevated inflammation (henceforth referred to as "tau pathology/degeneration" (PC1)). The second PC loads specifically on Aβ42 only ("Aβ Pathology" PC2), a very early and important marker of amyloid deposition in the brain [2]. The third PC loads strongly on neurofilament light chain (NfL), but also moderately on YKL-40,

and can be interpreted as representing neuronal injury and the accompanying inflammatory response ("injury/inflammation" PC3), as NfL is a component of axons and its presence in CSF is a non-specific marker of neuronal damage [6]. The fourth PC loads on YKL-40 and only weakly on tau and NfL, and therefore can be regarded as representing neuronal inflammation and astroglial reaction, not related to AD symptoms ("non-AD inflammation" PC4). Similarly, the fifth PC loads strongly on Ng and weakly on tau, representing synaptic functioning mostly independent of the other biomarkers and AD symptoms ("non-AD synaptic functioning" PC5). Please note, that the PC "names" are used for increased legibility, but are by necessity reductionist. We do not wish to imply that the named biological processes are unique or exhaustively describe any given PC.

After establishing the component structure, we applied these to search for rare variant associations using whole-exome sequencing in our previous study [3]. This work led to the identification of six genes, in which rare variants were associated with the CSF PCs. Specifically, we identified associations between the injury/inflammation component (PC3) and rare variants in *IFFO1*, *DTNB*, *NLRC3*, and *SLC22A10*, as well as between the non-AD synaptic functioning component (PC5) and rare variants in *GABBR2* and *CASZ* [3]. Interestingly, rare variant associations with AD risk were simultaneously reported for the *DTNB*locus in an independent project utilizing whole-genome sequencing in AD families and case-control datasets [11].

In this study, we aimed to extend the previous, rare-variant analyses to investigate the role of common variants on the PCA-defined CSF biomarker profiles. While previous GWAS in the field have screened for common-variant associations with single biomarkers [9, 12–15], to our knowledge no GWAS combining these CSF biomarkers in a multivariate framework has been performed to date. Multivariate analyses have the advantages of (i) allowing a more robust (compared to univariate analyses) quantification of different disease pathways, resulting in increased statistical power [16, 17] and (ii) enabling to differentiate various possible mechanisms of action more precisely.

Secondary aims of our study included the identification of sex-specific effects and AD mediation pathways. AD is more prevalent in women, and CSF biomarkers

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 3 of 23

differentially predict brain and cognitive changes depending on sex [18, 19]. Furthermore, genetic effects on CSF biomarkers may depend on sex as well, e.g., rs34331204 on chromosome 7p21 was found to have a male-specific association with neurofibrillary tangles [20]. It is therefore prudent to investigate whether the component structure differs between sexes and whether associations of PCs with AD or with genetic predictors is sex-dependent. Finally, we performed mediation analyses to gauge whether potential SNP effects on CSF biomarker profiles also affect AD risk.

Methods

Participants

The presented work is part of the EMIF-AD project, a consortium of European studies investigating the etiology of AD and AD biomarkers with the aim to improve prognosis and diagnosis [4]. Participants included elderly individuals with cognitively unimpaired individuals, mild cognitive impairment (MCI), and AD type dementia. Both deep phenotyping (such as brain imaging and determination of CSF biomarkers) and genotyping (SNP arrays and WES) were performed on a large number of EMIF-AD participants [21–23]. The current study utilizes the existing CSF biomarker and SNP array data and combines them with a range of statistical methods not previously employed on these data. Written informed consent was obtained for all assessment before the start of the study [4]. The study was conducted in accordance to the Declaration of Helsinki and ethical approval was obtained from the Ethical Committee of the University of Lübeck, as well as local committees of consortium members [4]. More details on the recruitment and phenotype ascertainment protocols used in the EMIF-AD dataset can be found in Bos et al. [4].

To increase the generalizability of effect estimates and to increase power to detect new associations, we performed all analyses jointly with equivalent CSF biomarker and SNP genotype data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [5]. Data used in the preparation of this article were obtained from adni.loni.usc.edu. ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD.

The current study utilized two participant selection paradigms for analysis: first, we selected participants for whom observations for at least 4 out of the 5 biomarkers were available. In total, this yielded 1158 participants to construct and examine the biomarker

PC scores (Additional file 1: Table S1). Second, we only included participants with available SNP array data, who were unrelated and of European ancestry. This reduced the sample size to 973 participants (Additional file 1: Table S2, see also Hong et al. for detailed selection methods [13]). Overall, both EMIF-AD and ADNI were comparable datasets of elderly participants, with a mean age at ascertainment of 69 and 75 years, respectively (Additional file 1: Table S1). The distributions of diagnostic status were similar in both datasets as well, with approximately half of the sample diagnosed with MCI, while 25% presented either no cognitive impairment or with a diagnosis of AD (Additional file 1: Table S1).

We further assessed novel findings in the independent Knight Alzheimer Disease Research Center (Knight-ADRC) cohort. Applying the same inclusion criteria, we selected 786 participants with sufficient genotype and phenotype/CSF data. These were on average 69 years old and 46.95% male, 21.12% had MCI, and 4.71% AD (Additional file 1: Table S3, Additional file 2: Supplementary Methods).

Measures

Genotyping, imputation, and quality control (QC)

SNP genotypes were determined using the Infinium Global Screening Array (GSA; Illumina, Inc., USA) at the Institute of Clinical Molecular Biology (UKSH, Campus-Kiel) in EMIF-AD and using Illumina's Omni 2.5 M or Human610-Quad arrays in ADNI. Autosomal SNPs in both GWAS datasets were processed with the same computational workflow [13], including the imputation of untyped variants with MiniMac 3 using the HRC 1.1 reference panel [24]. Here, we only analyzed common SNPs (MAF ≥ 0.01 per study) with sufficient imputation quality ($R^2 > 0.30$) and SNPs within HWE ($p < 5 \times 10^{-6}$). To check for cryptic relatedness, we used the KING-robust kinship estimator implemented in PLINK (v2.0) with a cutoff of 0.025, which excludes family relationships down to the fourth degree. Please see Hong et al. for a detailed description of the GWAS methods, QC criteria, and processing pipeline [13]. Within Knight-ADRC, participants were genotyped using GSA, CoreExome, Illumina 660 K, NeuroX2, and OmniExpress arrays. All samples were imputed using the TOPMed imputation server and only SNPs with an imputation quality of $R^2 > 0.30$ and HWE ($P \geq 1*10^{-6}$) were kept. Related individuals were identified through identity by descent (IBD; PLINK (v1.9)) with PI_HAT > 0.20 cutoff. For analysis, only samples of European ancestry, as determined by genotype PCs, and SNPs with MAF ≥ 0.0001 (0.01%) and 98% call rate were kept. For X-chromosome-specific methods, see Additional file 2: Supplementary Methods.

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 4 of 23

CSF biomarkers

Biomarkers were derived from CSF, as obtained via lumbar puncture [3, 5, 23]. For EMIF, the V-PLEX Plus AbPeptidePanel 1 Kit was used to measure Aβ, and in the case of tau, the INNOTEST ELISA was applied [23]. In ADNI, the Elecsys CSF immunoassay and a cobas e 601 analyzer assessed Aβ and tau concentration [25]. For both cohorts, ELISA was applied to assess NfL levels [23, 26]. Ng concentration was measured by ELISA in EMIF [23] and by electrochemiluminescence in ADNI [27]. ELISA was used to measure YKL-40 levels in EMIF [23] and LC/MRM-MS proteomics were applied in ADNI [28]. For the YKL-40 proteomics data, we *z*-score standardized two ion frequencies with two peptide sequences each and averaged the values. In Knight-ADRC, Aβ, tau, and ptau levels were measured using Lumipulse G1200 automated assay system. NfL, Ng, and YKL-40 levels were obtained using the aptamer-based SOMAscan (v4) platform.

Statistical analysis

CSF biomarker PCA, sex differences, and AD associations

The analysis work-flow is summarized in Additional file 2: Fig. S1. First, we computed five PCs across all participants with sufficient biomarker information. PCs were defined as described previously [3] and assigned to specific functional domains, as described in the introduction. The PCA was performed on the phenotype level and PCs were constructed independent of genotype information. Briefly, we first applied a rank-based inverse normal transformation within both studies to decrease extreme skewness of the observed biomarker levels and to *z*-score standardize their scale across studies [29]. We used the missMDA package to determine the optimum number of components and account for missing data [30]. Specifically, we applied leave-one-out cross-validation, removing one observation at a time and predicting it by a PCA model fitted to the rest of the dataset. The model resulting in the smallest mean square error of prediction contained five components (PC1-PC5). We then imputed missing values to avoid excessive sample size losses and potential participation biases. This was achieved using a regularized iterative PCA method with five components, as implemented in missMDA. Finally, we performed a PCA with varimax rotation (an orthogonal rotation) and extracted five PC scores with the psych package [31]. All analyses were performed in R 4.0.3 [32] We applied the same PC loadings to the Knight-ADRC for replication purposes of results related to PC4 and PC5.

We have not previously explored to which degree the component structure differs between sexes, or whether the resulting PCs show sex-dependent associations with dementia symptoms. We first repeated PCA in both males and females and compared loadings. We then tested whether mean PC levels differed between sexes. This was achieved by regressing PC scores on sex, adjusted for age, five genetic ancestry components, diagnostic status (dummy coding MCI and AD), and study (ADNI vs EMIF-AD).

In a last step, we used the PC scores as predictor of latent AD. Here, latent AD is defined as an underlying continuous normally distributed variable, representing a range of probability to either have no cognitive impairment, MCI, or AD. Latent AD was estimated by item factor analysis [33]. Accordingly, participants with low scores (below −1.47SD) are unlikely to display cognitive impairment, above −1.47SD and below 0.40SD are most likely to suffer from MCI and above 0.40SD have a high probability to be affected by AD. To account for potential sex differences, we also added a product term between PC scores and sex, coded as male=0 and female=1. The biomarker PC term can therefore be interpreted as the association of biomarker PC scores on latent AD in males and the interaction term as the female-specific effect, i.e., the difference between sexes. These analyses were adjusted for the same covariates as in the main analyses, i.e., age, genetic ancestry, and study. We applied a structural equation model (SEM) with a weighted least square mean and variance adjusted (WLSMV) estimator using Lavaan 0.6–9 [34] to estimate latent AD and regress it onto biomarker PCs, sex, their interaction, and covariates.

GWAS and meta-analyses

We performed four sets of GWAS: main-effect GWAS analysis (both sexes), male-only GWAS, female-only GWAS, and sex interaction analyses. Within each analysis group, we performed five GWAS, one for each biomarker PC, separately for EMIF-AD and ADNI. For all GWAS, separate linear regression models were run in PLINK [35] to all autosomal and X-linked SNPs passing QC. For the X-chromosomal analyses genotype, dosage for hemizygous males was coded as 2, to reflect the same dosage as homozygous females [36]. The biomarker PC scores were treated as outcome, imputed SNP dosage (0–2 numbers of effect allele) as main predictor. In addition, we included sex and five PCs reflecting genetic ancestry as additional covariates in the regression models. Analyses of the ADNI dataset were additionally corrected for genotyping array. Lastly, GWAS results for the EMIF-AD and ADNI were meta-analyzed using the inverse variance weighting (i.e., fixed-effect) model implemented in METAL [37].

In secondary analyses, we aimed to discover SNP effects exclusively found in one sex by running sex-specific GWAS. In the final model, we added a sex interaction term, representing the difference between the SNP

effect in females vs males. In addition to single-variant analyses, we also estimated the aggregate effect of all SNPs within a protein-coding gene. These analyses were performed with MAGMA 1.08 [38] on the FUMA 1.3.7 platform with default settings [39]. FUMA was also used to select independent ($R^2 \leq 0.6$) genome-wide significant ($p < 5*10^{-8}$) SNPs for further mediation analyses and to perform gene-set enrichment analyses on the GWAS top hits emerging for each analyzed PC. While the focus was on multivariate analyses using PCs as outcome, we also performed the same analyses with single biomarkers to compare both approaches.

We further estimated the variance explained by loci found to be genome-wide significant in the main-effect analyses. We computed the difference in R^2 between a full model including the tested SNP and a baseline model with covariates only. We additionally estimated the local heritability of the top loci using LAVA 0.1.0 [40]. We defined the region of interest as 10 KB down or upstream

of the gene start/end or SNPs, or in the case of *APOE*, as previously defined [41] (Additional file 1: Table S4). Single-SNP, gene-based and local heritability analyses were repeated in Knight-ADRC for the novel PC4 and PC5 loci showing genome-wide significance in the main discovery analyses to assess replication.

Mediation analyses

Independent SNPs, which showed genome-wide significant association in any of the GWAS were further tested for mediation effects. Specifically, we examined whether these SNPs would affect latent AD via their influence on biomarker levels. To test this hypothesis, we applied a SEM to each SNP. In this SEM, the genetic variant predicts all biomarker PCs, as well as latent AD directly. The biomarker PCs in turn also predict latent AD. See Fig. 1 for a path diagram. Sex, age, five genetic ancestry components, and study were predictors of both CSF biomarkers and latent AD, thus all mediation and direct
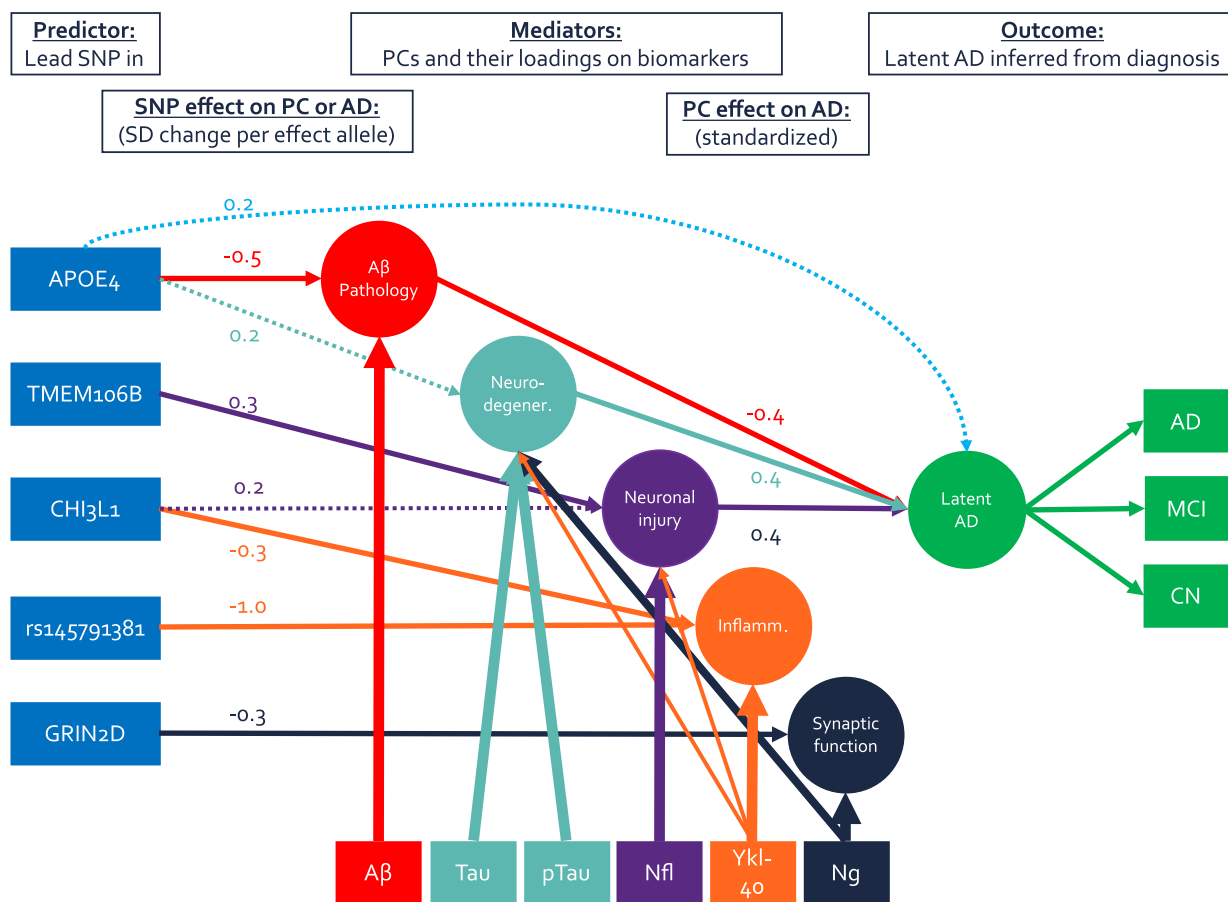

Fig. 1 Path model of main findings. This path model summarizes the main-effect mediation model. Circles indicate principal components or latent variables, rectangles represent observed variables. Arrows either indicate PC loadings or structural regression paths. Thicker lines correspond to stronger loadings, solid structural paths are genome-wide significant ($p < 5 \times 10^{-8}$), and dashed lines are suggestive ($p < 0.008$). Coefficients indicate either the effect of one effect allele on a biomarker PC in SD, or the effect of one SD higher biomarker PC score on latent AD in SD. Note: all paths are adjusted for assessment age, sex, genetic ancestry, and study but are omitted from figure

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 6 of 23

pathways were statistically adjusted for these potential confounders. Paths from SNPs to a biomarker were multiplied with the path from biomarker to latent AD to obtain the mediation effect for that particular PC. We also summed all mediation pathways to obtain a total mediation estimate elicited by any biomarker. This biomarker mediation estimate was further summed with the direct effect to estimate the total effect. To aid the interpretation of the mediation magnitude, we also provide an estimate of the proportion of the mediated effect (i.e., biomarker mediation/total effect). However, this was only possible, when mediated and total effects pointed in the same direction. Additional context to the total effect is afforded by providing the variants' effects on AD based on a large and independent previous case-control GWAS [42].

As some SNPs showed sex-dependent effects, we also ran a moderated mediation model to account for sex-specific mediation. This was achieved by adding a product term between SNP dosage and dummy variable for sex (female = 1, male = 0), and adding this interaction term as predictor of biomarker PCs and latent AD. If no mediation pathway differed nominally ($p \geq 0.05$) between sexes, main mediation model results are presented. Otherwise, male- and female-specific mediation estimates are provided, as estimated by the moderated mediation model. All mediation analyses were estimated with WLSMV in lavaan [34].

Comparison to rare variant results

As outlined in the introduction, we previously identified several rare variant associations using the same biomarker PCA approach in a subset of the EMIF-AD individuals analyzed here. Specifically, this pertains to associations between *IFFO1*, *DTNB*, *NLRC3*, and *SLC22A10* and the injury/inflammation component (PC3), as well as between *GABBR2* and *CASZ* and the non-AD synaptic functioning component (PC5) [3]. Here we examined, whether common variants—in addition to the rare variants already identified—in these genes also show associations with the CSF biomarker PCs, using both single-variant and gene-based tests as outlined above.

Functional characterization and related phenotypes

SNPs were annotated to genes using default settings of FUMA, which relies on positional and functional information. Given the broad association signal between the chromosomal region encompassing *GRIN2D* and non-AD synaptic functioning (PC5), we additionally performed fine-mapping analyses to identify the most likely causal genes at this locus. To this end, we used a TWAS

fine-mapping approach [43] as implemented in FOCUS 0.802 (https://github.com/bogdanlab/focus).

Briefly, we examined the association between predicted gene expression at the *GRIN2D* locus with non-AD synaptic functioning (PC5). The *GRIN2D* locus was defined as encompassing chr19:48382575–48466980 (GRCh38/hg38), starting 10 kb downstream of *KDELR1* and 10 kb upstream of *KCNJ14* to cover all lead variants within the locus. Prediction weights were previously computed based on expression in the dorsolateral prefrontal cortex of ROS/MAP participants with AD, MCI, or healthy controls as obtained from the AMP-AD RNAseq Harmonization Study [44, 45], and processed by Bellenguez et al. [46] We chose this dataset as it includes a relatively large number of samples of a relevant tissue in a population similar to the current study. Prediction quality of *GRIN2D* was sufficient in this dataset, based on significant variance explained as estimated with cross-validation (5.2%, $p = 3.6 \times 10^{-8}$). In total, weights for 13,420 transcripts with nominally significant heritability estimates were available. We next associated the predicted transcripts with non-AD synaptic functioning (PC5) and consequently computed the posterior inclusion probability (PIP), resulting in a credible gene set. We only considered genes as likely causal candidates, if they showed an association with the outcome on a transcriptome-wide significance level ($p < 0.05/13,420 < 3.7 \times 10^{-6}$; $|Z| > 4.6$).

To further study the impact of *GRIN2D* on related phenotypes, we tested the predicted gene expression for association with risk for AD [42, 46], low educational attainment [47], low cognitive ability [48], and risk for major depressive disorder [49] using summary statistics of independent, recent GWAS. Finally, we also attempted to compute local genetic correlations between PCs and related phenotypes. However, due to low sample sizes, this was only possible for *TMEM106B* and the genetic correlation between injury/inflammation (PC3) and AD. Similarly, we were not able to investigate genetic correlations between sexes given the even lower sample sizes in these subsets.

Multiple testing adjustment

To strike a balance between reliable inference and power, we present our findings as primary, secondary, and tertiary results. The primary analyses in this study were the GWAS in the full dataset independent of sex. For SNP-based tests, we apply the conventional genome-wide association threshold of $p < 5 \times 10^{-8}$, and for gene-based tests, we used Bonferroni's method to adjust for 19,511 genes resulting in a threshold of $p < 2.3 \times 10^{-6}$, as recommended by FUMA. The sex-specific analyses present additional tests of related (and non-independent) hypotheses, and, thus should be regarded as secondary

Table 1 PCA loadings. Component loadings of each biomarker (first column) on five principal components (column groups two to six) are displayed

Sample	Tau pathology/Degeneration (PC1)			Aβ Pathology (PC2)			Injury/Inflammation (PC3)			Non-AD Inflammation (PC4)			Non-AD Synaptic functioning (PC5)		
	Male (n=601)	Female (n=557)	All (n=1158)	Male (n=601)	Female (n=557)	All (n=1158)	Male (n=601)	Female (n=557)	All (n=1158)	Male (n=601)	Female (n=557)	All (n=1158)	Male (n=601)	Female (n=557)	All (n=1158)
Tau	0.86	0.87	0.87	−0.05	−0.07	−0.06	0.23	0.25	0.23	0.26	0.24	0.24	0.27	0.27	0.27
pTau	0.91	0.89	0.91	−0.08	−0.07	−0.08	0.11	0.18	0.14	0.19	0.23	0.20	0.26	0.28	0.26
Aβ	−0.07	−0.07	−0.07	1.00	1.00	1.00	−0.03	−0.02	−0.03	−0.01	0.00	0.00	0.04	0.04	0.04
NfL	0.19	0.25	0.21	−0.04	−0.03	−0.03	0.95	0.93	0.94	0.25	0.27	0.25	0.08	0.10	0.08
YKL-40	0.31	0.33	0.32	−0.01	0.01	0.00	0.30	0.33	0.32	0.88	0.87	0.88	0.17	0.17	0.17
Ng	0.49	0.51	0.51	0.06	0.07	0.07	0.11	0.12	0.10	0.18	0.19	0.18	0.84	0.83	0.83
R^2	0.33	0.33	0.33	0.17	0.17	0.17	0.18	0.18	0.18	0.16	0.16	0.16	0.15	0.15	0.15

R^2 Variance explained by PC

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 8 of 23

and more exploratory analyses. For the mediation analyses, we applied an alpha of $0.05/6 = p < 0.0083$, adjusting for six potential mediation or direct pathways. See Additional file 2: Supplementary Methods for full description of the multiple testing adjustment strategy.

Results

CSF biomarker PCs

Analogous to our previous study [3], the six AD CSF biomarkers tested here could be combined into five consistent components across datasets and analytical subsamples. The PC structure was very similar across the EMIF-AD and ADNI datasets (Additional file 1: Table S5); therefore, all subsequent PC analyses were based on a combined discovery sample to maximize sample size and reduce study heterogeneity. In this study, we extended our analyses to examine whether the CSF biomarker PCs' loadings, their mean levels, or associations with latent AD differ by sex. Generally, PC loadings were consistent across males and females (Table 1). NfL loaded 0.06 higher on tau pathology/degeneration in females when compared to males (0.25 vs 0.19). A similar pattern was observed for pTau, which loaded 0.07 higher on Injury/Inflammation (0.18 vs 0.11) in women vs men. All other loading differences were below 0.04 and therefore classified as "indifferent" between sexes. Based on these observations, we used the common loadings as estimated across both sexes for further analyses.

While the component structure was very similar between sexes in general, we observed differences in mean levels. When adjusting for age, diagnostic status, and study, females showed 0.21SD (SE = 0.06, $p = 6.3 \times 10^{-4}$) higher scores on non-AD synaptic functioning. In contrast, injury/inflammation was −0.40SD (SE = 0.05, $p = 2.7 \times 10^{-13}$) lower in females. Tau pathology/degeneration, Aβ pathology, and non-AD inflammation did not show differences in mean levels across sexes when accounting for multiple testing ($p < 0.05/5 = 0.01$).

Finally, we examined the association of the biomarker PCs with latent AD, including potential sex interactions (Additional file 1: Table S6). 1SD higher levels in tau pathology/degeneration or injury/inflammation were associated with 0.41SD and 0.40SD higher latent AD levels in males. Females had a stronger association with 0.43SD and 0.44SD, respectively, but the difference was not significant ($p > 0.5$). Higher brain Aβ accumulation is reflected in lower CSF Aβ values; therefore, higher Aβ pathology scores were associated with lower AD occurrence in both males and females ($\beta = -0.34$SD and $\beta = -0.45$), although the difference between sexes was not significant ($p = 0.13$). Non-AD inflammation and non-AD synaptic functioning did not significantly

associate with latent AD, when adjusting for five tests (i.e., all $p > 0.01$). Due to lack of evidence for sex-differential associations, all subsequent analyses are performed under the assumption that associations between PCs and latent AD are invariant across sexes.

GWAS

In our GWAS analyses, we tested 7,433,949 autosomal and X-linked SNPs. For gene-based tests, we assessed 19,511 protein-coding genes. For all outcomes and analyses, lambda was below 1.05 and QQ-plots showed no evidence of noteworthy genome-wide inflation (Additional file 2: Fig. S2). GWAS results are visualized as Manhattan plots for main (Fig. 2), sex stratified (Fig. 3), sex interaction analyses (Additional file 2: Fig. S3), and gene-based tests (Additional file 2: Fig. S4). Results of independent SNPs showing genome-wide significance are summarized in Table 2. Their effect on AD risk by mediation analysis using CSF biomarker PCs are displayed in Table 3. Finally, gene-based results are depicted in Table 4. We had good to excellent power to detect SNPs with a β of 0.30 and a minor allele frequency (MAF) of at least 10% [50]. For less common SNPs (< 5%), we only had power to detect SNPs with a β of 0.65 or higher (Additional file 2: Fig. S5).

Main analyses

In the main analyses, none of the SNPs showed genome-wide significant association with the tau pathology/degeneration component (PC1). In contrast, seven independent SNPs in the *APOE* locus showed genome-wide significant associations with the Aβ pathology component (PC2). Specifically, the C allele of the lead SNP rs429358 (which is also known as the ε4-allele) was associated with −0.50SD lower PC scores (SE = 0.04, $p = 1.3 \times 10^{-29}$, MAF = 0.30).

Regarding the injury/inflammation component (PC3), transmembrane protein 106B (*TMEM106B*), tagged by lead intronic SNP rs2302634, showed strong and genome-wide significant associations: the T allele at this variant predicted +0.26SD (SE = 0.04, $p = 1.3 \times 10^{-9}$, MAF = 0.42) higher injury/inflammation scores.

In case of the non-AD inflammation component (PC4), two loci reached genome-wide significance: chitinase 3 like 1 (*CHI3L1)* on chromosome 1 and a new region on chromosome 9p21.3 with lead SNP rs145791381 (located in an intergenic region). *CHI3L1* encodes the YKL-40 protein and two independent SNPs in or near this gene showed genome-wide significant associations with the non-AD inflammation PC: the strongest effect was observed for intronic variant rs7551263 (T allele: $\beta = -0.39$SD, SE = 0.05, $p = 5.7 \times 10^{-17}$, MAF = 0.16;

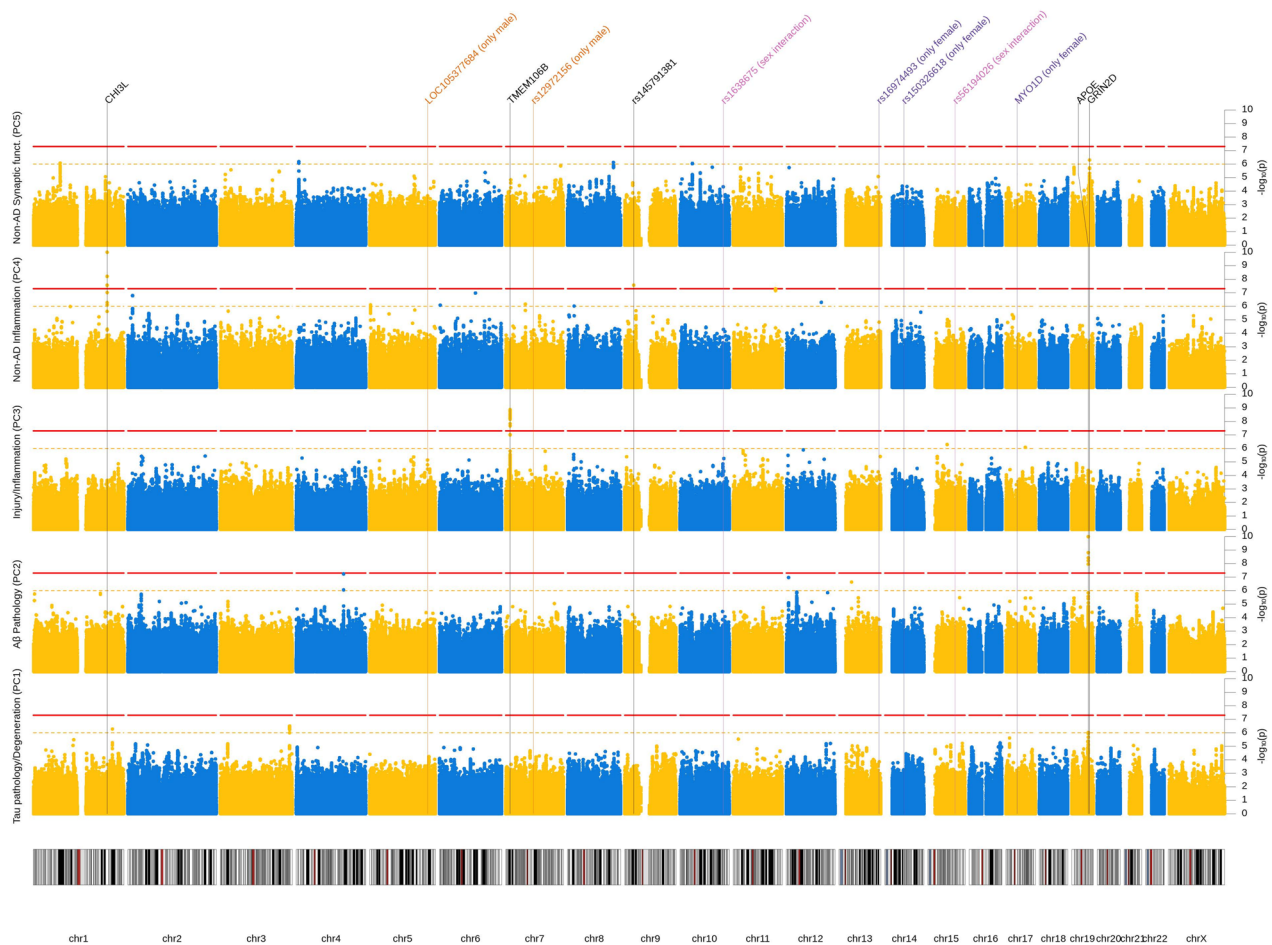Neumann *et al. Genome Medicine*   (2023) 15:79

Page 9 of 23

Fig. 2 Manhattan plot (main-effect model). Results from GWAS of five CSF biomarker PC across both sexes. Each row represents a different PC as outcome. *X*-axis represents each chromosome and the *y*-axis the *p*-value of the SNP association with the outcome on a $-\log_{10}$ scale. All analyses were adjusted for sex, genetic ancestry, and SNP array. Red line indicates genome-wide significance threshold ($p = 5 \times 10^{-8}$). Yellow line indicates suggestive threshold ($p = 1 \times 10^{-6}$). Vertical lines point towards genome-wide significant loci based on any model. *P*-values below 1×10^{-10} were winsorized to 1×10^{-10}

Fig. 4). The second signal in *CHI3L1* was elicited by SNP rs10399931 located < 160 bp upstream of *CHI3L1* (T allele: $\beta = -0.33$SD, SE = 0.06, $p = 6.0 \times 10^{-9}$, MAF = 0.24). Both SNPs are independent ($r^2 = 0.04$ in our study and D' = 0.71 in European population [39]) and therefore probably represent two separate signals. In a previous work [13], both SNPs were directly associated with CSF YKL-40 levels and can therefore be

considered *cis* pQTLs (i.e., protein quantitative trait loci) of this protein.

Another SNP reaching genome-wide significance in the main-effect analyses was the intergenic SNP rs145791381 on chr. 9p21.3. The T allele of this variant was associated with lower scores on the non-AD inflammation component (PC4, $\beta = -1.01$SD, SE = 0.18, $p = 6.0 \times 10^{-9}$, MAF = 0.02). Despite the strength of the association in

(See figure on next page.)

Fig. 3 Manhattan plot (sex stratified). Results from GWAS of five CSF biomarker PC for males and females separately. Each row represents a different PC as outcome. Per outcome, results for males are depicted at the bottom and for females at the top. *X*-axis represents each chromosome and the *y*-axis the *p*-value of the SNP association with the outcome on a $-\log_{10}$ scale. All analyses were adjusted for genetic ancestry and SNP array. Red line indicates genome-wide significance threshold ($p = 5 \times 10^{-8}$). Yellow line indicates suggestive threshold ($p = 1 \times 10^{-6}$). Vertical lines point towards genome-wide significant loci based on any model. *P*-values below 1×10^{-10} were winsorized to 1×10^{-10}

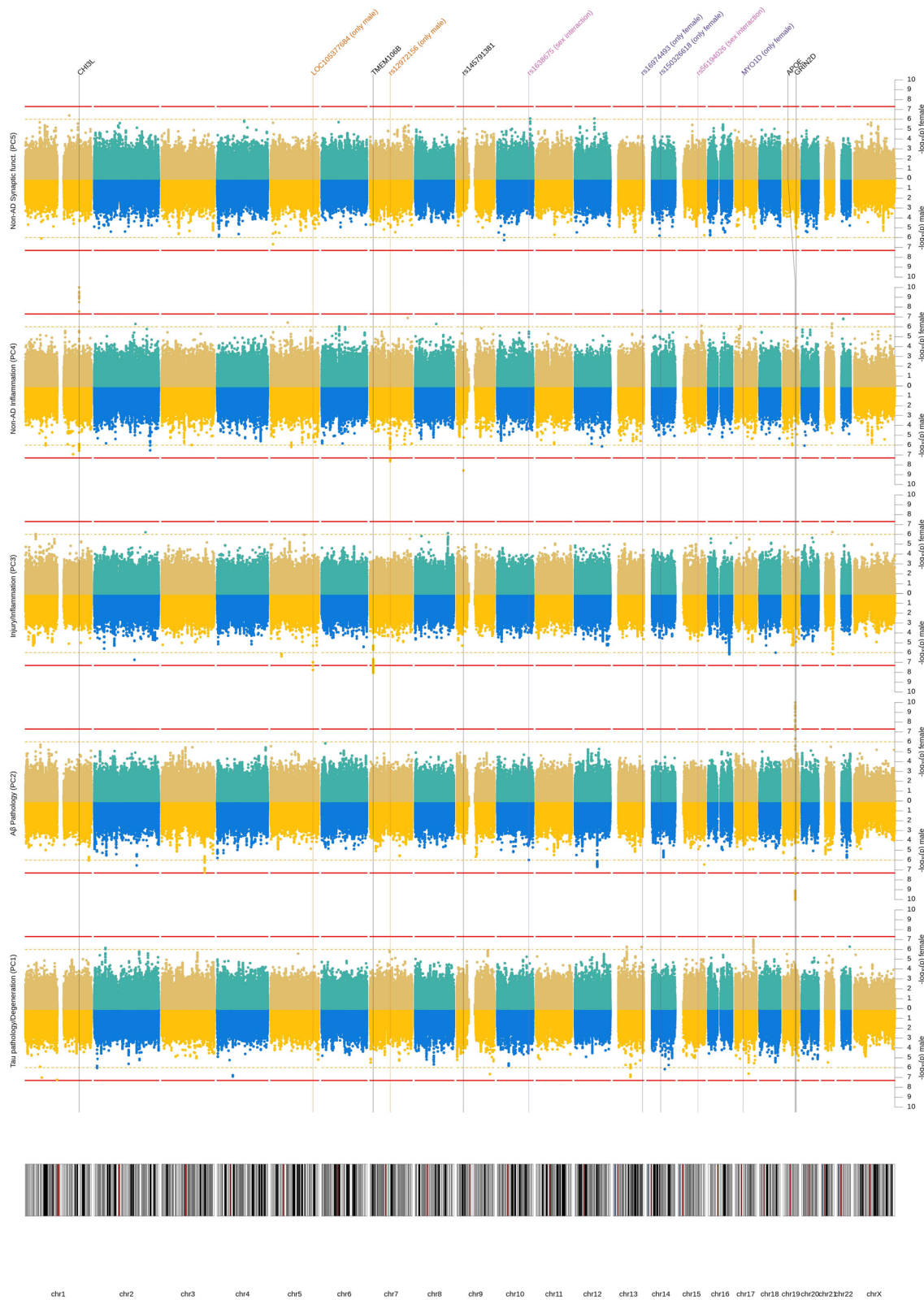Neumann *et al. Genome Medicine*   (2023) 15:79

Page 10 of 23

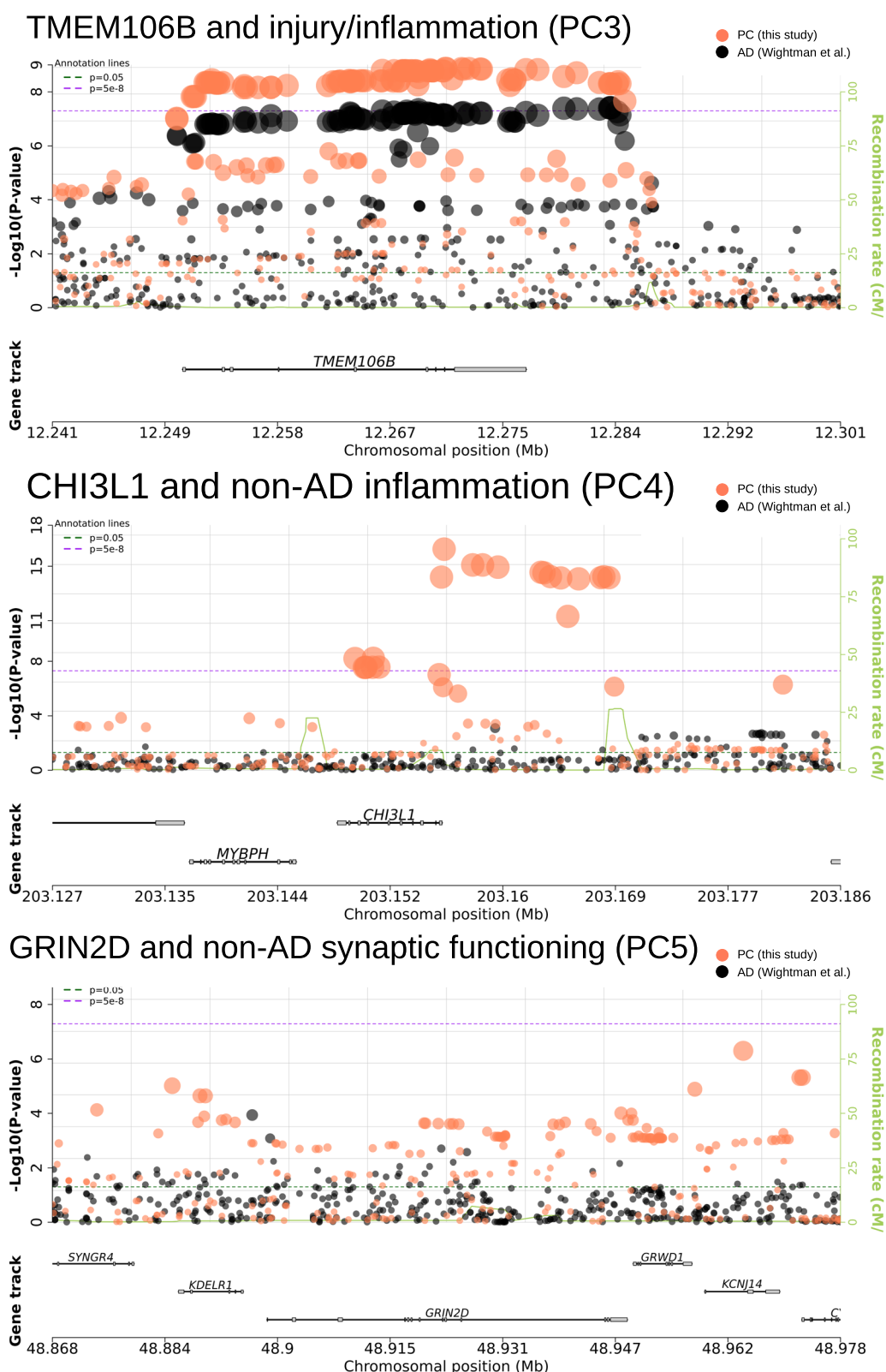**Fig. 3** (See legend on previous page.)

Fig. 4 Regional plots for *TMEM106B*, *CHI3L1*, and *GRIN2D*. Each plot displays the *p*-values of SNP associations in either *TMEM106B*, *CHI3L1*, or *GRIN2D* loci. Statistics are derived from two studies. Orange dots represent *p*-values of association with biomarker PCs estimated in this study and back dots represent *p*-values of association with AD, as estimated in a separate GWAS on 1 million participants (Wightman et al. [42]). Regional plots were created with snpxplorer.net

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 12 of 23

Table 2 Independent genome-wide significant loci. Independent SNPs, which had genome-wide significant associations with a Biomarker PC in either main effect, sex-stratified or sex interaction models. See also Supplementary Tables for fine-mapping and replication analyses of the PC5 signal near *GRIN2D*

SNP	Chr:Pos	Location	Locus	EA	EAF	R^2_{imp}	Main effect			Male Stratified			Female Stratified			Sex Interaction		
							β	SE	p	β	SE	p	β	SE	p	β	SE	p
Tau pathology/Degeneration (PC1)																		
rs140169162	17:30886881	Intron	*MYO1D*	G	0.98	0.69	0.64	0.19	8.6E−04	−0.02	0.26	9.4E−01	1.61	0.30	4.9E−08	0.84	0.23	3.0E−04
Aβ Pathology (PC2)																		
rs1638675	10:119178376	Intergenic	-	A	0.57	0.89	−0.08	0.05	7.1E−02	−0.33	0.07	1.0E−06	0.18	0.06	4.2E−03	0.50	0.09	3.5E−08
rs12972156	19:45387459	Intron	*APOE*	C	0.74	0.98	0.43	0.05	9.5E−20	0.46	0.07	6.9E−11	0.39	0.07	2.3E−09	−0.04	0.09	6.5E−01
rs157582	19:45396219	Intron	*APOE*	C	0.65	1.00	0.37	0.04	9.5E−18	0.40	0.06	2.6E−10	0.33	0.06	3.5E−08	−0.05	0.09	5.6E−01
rs10119	19:45406673	3'-UTR	*APOE*	G	0.60	0.95	0.34	0.04	6.4E−15	0.43	0.06	4.9E−11	0.29	0.06	2.6E−06	−0.12	0.09	1.5E−01
rs429358	19:45411941	Exon	*APOE*	T	0.70	0.99	0.50	0.04	1.3E−29	0.55	0.06	1.2E−17	0.44	0.06	4.3E−12	−0.09	0.09	3.0E−01
rs75627662	19:45413576	Intergenic	*APOE*	C	0.71	0.98	0.35	0.05	1.2E−13	0.38	0.07	4.4E−08	0.31	0.07	2.7E−06	−0.05	0.09	6.2E−01
rs157595	19:45425460	Intergenic	*APOE*	A	0.33	0.96	0.28	0.05	1.6E−09	0.28	0.07	7.1E−05	0.24	0.06	2.3E−04	0.00	0.09	9.8E−01
rs111789331	19:45427125	Intergenic	*APOE*	T	0.73	0.95	0.45	0.05	8.3E−21	0.46	0.07	1.7E−10	0.45	0.07	1.6E−11	0.02	0.09	8.0E−01
Injury/Inflammation (PC3)																		
rs114211800	5:158973225	Intron	*LOC105377684*	G	0.98	0.91	−0.74	0.17	7.2E−06	−1.23	0.22	1.7E−08	−0.11	0.26	6.9E−01	1.07	0.34	1.5E−03
rs2302634	7:12270770	Intron	*TMEM106B*	T	0.58	1.00	0.26	0.04	1.3E−09	0.34	0.06	3.2E−08	0.18	0.06	2.9E−03	−0.16	0.09	6.2E−02
rs56194026	15:74831534	Intergenic	*ARID3B*	C	0.92	0.90	−0.03	0.08	7.2E−01	−0.50	0.12	1.6E−05	0.41	0.12	5.3E−04	0.90	0.16	4.7E−08
Non-AD Inflammation (PC4)																		
rs7551263	1:203150756	Intron	*CHI3L1*	T	0.84	0.99	−0.33	0.06	6.0E−09	−0.30	0.08	8.7E−05	−0.40	0.09	3.1E−06	0.07	0.12	5.5E−01
rs10399931	1:203156080	Intergenic	*CHI3L1*	T	0.24	0.98	−0.39	0.05	5.7E−17	−0.36	0.07	2.7E−07	−0.45	0.07	1.0E−11	−0.12	0.09	1.9E−01
rs12670437	7:76390507	Intergenic	-	G	0.98	0.87	−0.49	0.15	9.9E−04	−1.20	0.22	2.4E−08	0.20	0.21	3.3E−01	1.24	0.32	1.2E−04
rs145791381	9:24796757	Intergenic	-	T	0.98	0.93	−1.01	0.18	2.7E−08	−1.27	0.21	2.7E−09	−0.41	0.36	2.6E−01	NA	NA	NA
rs6974493	13:110307768	Intergenic	-	T	0.97	0.98	0.59	0.14	1.5E−05	0.18	0.22	4.2E−01	0.97	0.17	2.2E−08	0.77	0.27	5.2E−03
rs150326618	14:52641990	Intergenic	-	G	0.98	0.80	0.57	0.17	9.8E−04	0.09	0.24	7.2E−01	1.39	0.25	2.6E−08	1.13	0.35	1.3E−03
Non-AD Synaptic Functioning (PC5)																		
rs275844	19:48884949	Intergenic	*GRIN2D*	G	0.88	0.93	−0.30	0.07	9.6E−06	−0.39	0.10	4.0E−05	−0.20	0.10	4.2E−02	0.22	0.14	1.1E−01
rs8111684	19:48947326	3'-UTR	*GRIN2D*	G	0.26	1.00	0.19	0.05	9.8E−05	0.24	0.07	8.4E−04	0.12	0.07	7.9E−02	0.11	0.10	2.9E−01
rs3107911	19:48964305	Intron	*GRIN2D*	C	0.15	0.96	0.31	0.06	5.0E−07	0.38	0.09	8.7E−06	0.21	0.09	2.0E−02	0.17	0.13	1.7E−01

SNP RS ID, *Chr:Pos* Chromosome:Position Build 37, *EA* effect allele, *SE* standard error, R^2_{imp} Imputation Quality, *β* effect of one copy of the effect allele on biomarker PCs in SD. In case of sex interaction, the effect specific to females, *SE* standard error, *p* p-value, *NA* rs145791381 interaction models failed to converge due to too high multicollinearity

Table 3 Mediation models. Results from structural equation models testing the hypothesis, that SNP effects on CSF biomarker PCs would also affect AD risk

SNP	Chr:Pos	Location	Locus	EA	EAF	R^2_{imp}	Sex	PC1	PC2	PC3	PC4	PC5	PC	Prop	Direct	Total	Z_{AD} (Wightman et al.) [42]
Tau pathology/Degeneration (PC1)																	
rs140169162	17:30886881	Intron	MYO1D	G	0.98	0.69	m	0.02	0.10	−0.03	0.01	0.00	0.11	0.29	0.26	0.36	−1.94
							f	0.52*	−0.12	−0.04	−0.01	0.00	0.35*		−1.04*	−0.69	
Aβ Pathology (PC2)																	
rs1638675	10:119178376	Intergenic	-	A	0.57	0.89	m	0.03	0.10*	−0.03	0.00	0.00	0.10*		−0.06	0.04	−0.18
							f	0.02	−0.04*	0.00	0.00	0.00	−0.02		0.07	0.04	
Tau pathology/Degeneration & Aβ Pathology (PC1 & PC2)																	
rs12972156	19:45387459	Intron	APOE	C	0.74	0.98	all	−0.07*	−0.12*	0.02	0.00	0.00	−0.17*	0.65	−0.09	−0.27*	−Inf*
rs157582	19:45396219	Intron	APOE	C	0.65	1.00	all	−0.08*	−0.11*	0.01	0.00	0.00	−0.17*	0.79	−0.05	−0.22*	−Inf*
rs10119	19:45406673	3'-UTR	APOE	G	0.60	0.95	all	−0.06*	−0.10*	0.01	0.00	0.00	−0.15*	0.67	−0.07	−0.23*	−Inf*
rs429358	19:45411941	Exon	APOE	T	0.70	0.99	all	−0.09*	−0.13*	0.01	0.00	0.00	−0.21*	0.54	−0.18*	−0.39*	−Inf*
rs75627662	19:45413576	Intergenic	APOE	C	0.71	0.98	all	−0.06*	−0.10*	0.02	0.00	0.00	−0.14*	0.59	−0.10	−0.24*	−Inf*
rs157595	19:45425460	Intergenic	APOE	A	0.33	0.96	all	−0.06*	−0.09*	0.02	0.00	0.00	−0.13*	0.54	−0.11	−0.23*	−29.68*
rs111789331	19:45427125	Intergenic	APOE	T	0.73	0.95	all	−0.08*	−0.12*	0.02	0.00	0.01	−0.19*	0.57	−0.14	−0.33*	−Inf*
Injury/Inflammation (PC3)																	
rs7551263	1:203150756	Intron	CHI3L1	T	0.84	0.99	all	0.04	−0.01	0.05*	−0.02	0.00	0.06		−0.11	−0.05	1.21
rs114211800	5:158973225	Intron	LOC105377684	G	0.98	0.91	m	0.11	−0.08	−0.31*	0.00	0.00	−0.29*	0.54	0.27	−0.02	0.45
							f	−0.09	−0.05	0.03	0.01	0.00	−0.09		0.58	0.49	
rs2302634	7:12270770	Intron	TMEM106B	T	0.58	1.00	all	−0.02	−0.01	0.07*	0.00	0.00	0.04		−0.08	−0.04	5.38*
rs56194026	15:74831534	Intergenic	ARID3B	C	0.92	0.90	m	−0.03	0.01	−0.13*	0.01	0.00	−0.15*	0.19	0.09	−0.06	−0.03
							f	−0.01	−0.03	0.10*	0.00	0.00	0.06		−0.18	−0.12	−0.01
No evidence for mediation																	
rs10399931	1:203156080	Intergenic	CHI3L1	T	0.24	0.98	all	0.04	−0.03	0.02	−0.03	0.00	−0.01		0.10	0.10	0.58
rs12670437	7:76390507	Intergenic	-	G	0.98	0.87	all	−0.10	0.05	−0.01	−0.01	0.00	−0.07		−0.06	−0.14	0.22
rs145791381	9:24796757	Intergenic	-	T	0.98	0.93	all	0.00	−0.01	0.10	−0.03	0.00	0.06		−0.08	−0.02	2.20*
rs16974493	13:110307768	Intergenic	-	T	0.97	0.98	all	−0.01	0.01	0.01	0.02	0.00	0.03		−0.07	−0.03	0.34
rs150326618	14:52641990	Intergenic	-	G	0.98	0.80	all	−0.08	−0.02	−0.07	0.03	0.00	−0.15	0.19	−0.62	−0.77	−0.22
rs275844	19:48884949	Intergenic	GRIN2D	G	0.88	0.93	m	−0.07	0.02	−0.01	0.00	0.00	−0.06		0.18	0.12	−2.13*
							f	−0.02	0.05	0.02	−0.01	0.01	0.05		−0.18	−0.13	
rs8111684	19:48947326	3'-UTR	GRIN2D	G	0.26	1.00	all	0.02	0.00	−0.02	0.00	0.00	0.00	0.04	−0.01	−0.01	0.72
rs3107911	19:48964305	Intron	GRIN2D	C	0.15	0.96	all	0.02	0.00	−0.01	0.00	0.00	0.00	0.03	−0.05	−0.05	1.31

SNP RS ID, *Chr:Pos* Chromosome:Position Build 37, *EA* effect allele, *SE* standard error, R^2_{imp} Imputation Quality, *Sex* In case of significant differences in mediation pathways between males and females based on moderated mediation models, pathway estimates are given separately for males and females, *PC1-PC5* The effect of one copy of the effect allele on latent AD in SD via mediation of the biomarker PC, *PC1* Tau pathology/Degeneration, *PC2* Aβ Pathology, *PC3* Injury/Inflammation, *PC4* Non-AD Inflammation, *PC5* Non-AD Synaptic Functioning, *PC* Joint effect of all mediation pathways via any biomarker, *Prop.* Proportion of mediated effect (PC/Total), *Direct* The effect of one copy of the effect allele on latent AD in SD not mediated by any biomarker PC, *Total* The joint effect of one copy of the effect allele on latent AD in SD by either mediation or direct effects, Z_{AD} (*Wightman et al.*) [42] Z test statistic based on previous GWAS of AD

* *p* < 0.008

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 14 of 23

Table 4 Results of gene-based tests. Results from gene-based tests with MAGMA representing the joint effects of common SNPs within the named gene (NB: for PC1 no genome-wide significant gene-based results were observed)

Outcome	Gene	Chr	Start	End	N_{snps}	Z	P
Genome-wide significant genes							
Aβ Pathology (PC2)	*APOE*	19	45409011	45412650	6	7.91	1.3E−15
Injury/Inflammation (PC3)	*TMEM106B*	7	12250867	12282993	182	5.75	4.6E−09
Non-AD Inflammation (PC4)	*CHI3L1*	1	203148059	203155877	28	4.53	2.9E−06
Non-AD Synaptic Functioning (PC5)	*GRIN2D*	19	48898132	48948188	110	4.63	1.8E−06
Comparison with rare-variant hits							
Injury/Inflammation	*IFFO1*	12	6647541	6665239	46	0.50	0.307
Injury/Inflammation	*DTNB*	2	25600067	25896503	385	−1.11	0.867
Injury/Inflammation	*NLRC3*	-	-	-	-	-	-
Injury/Inflammation	*SLC22A10*	11	62905339	63137190	564	−1.79	0.963
Non-AD Synaptic Functioning	*GABBR2*	9	101050391	101471479	1445	1.41	0.079
Non-AD Synaptic Functioning	*CASZ1*	1	10696661	10856707	387	0.32	0.374

Chr chromosome, *Start/End* SNPs between start and end were considered, *Nsnps* number of SNPs included in test, *Z* Z test statistic, *p* p-value

the primary GWAS, this potential association signal was elicited by a "singleton variant" (i.e., there was no "trail" of correlated variants showing similar association evidence). The missing trail of correlated variants may be the result of this variant's low minor allele frequency which substantially reduces the number of variants in strong LD in this particular region. Given the fact that this variant was not associated with the non-AD inflammation PC4 in Knight-ADRC ($\beta=0.07$SD, SE$=0.35$, $p=0.84$), we have put no further emphasis on this finding throughout the remainder of the manuscript.

Lastly, no single SNP showed genome-wide significant association with the non-AD synaptic functioning component (PC5). However, this PC showed evidence for genome-wide significant association in the gene-based GWAS analyses highlighting the glutamate receptor gene *GRIN2D*, based on aggregated test statistics across 110 SNPs located between 19:48,898,132 and 19:48,948,188 (Z$=+4.63$, $p=1.8\times10^{-6}$) (Table 4). The lead SNP within this gene was rs8111684 in the 3' UTR region of the *GRIN2D* gene ($\beta=+0.19$SD, SE$=0.05$, $p=9.8\times10^{-5}$, MAF$=0.26$). Interestingly, the genomic regions flanking the gene both p-ter and q-ter showed more significant single SNP associations. As these variants are located outside of the gene, they were not considered in the gene-based tests (Fig. 4). Notwithstanding, we included two additional SNPs from this adjacent region for further characterization: rs275844 and rs3107911. The strongest single-SNP association in the locus was elicited by rs3107911, located in the intron of the gene *KCNJ14* ($\beta=-0.31$SD, SE$=0.06$, $p=5\times10^{-7}$, MAF$=0.15$) and approx. 16 kb q-ter of *GRIN2D* (Fig. 4). The lead variant explained 2.2% of the variance and the *GRIN2D* locus as a whole explained 3% ($p=0.009$) of the variance

in non-AD synaptic functioning (PC5). See Additional file 1: Table S4 for local heritability and estimates of the variance explained for all loci.

Comparison of the PCA-based GWAS results with GWAS analyses run on individual biomarkers revealed that the PCA phenotype led to a higher statistical support (i.e., smaller *P*-values) in about half (i.e., 10 out of 21) of the top hits of the main-effect analyses (Additional file 1: Table S7). This pattern was observed for 4 of the 5 computed PCs, i.e., all but PC2 (Aβ pathology). Only for this latter PC all ($n=8$) single biomarker results consistently showed stronger statistical support than the PCA-based GWAS. In contrast, comparing results for non-AD inflammation (PC4) vs. single biomarker analyses using YKL-40 levels, the associations tended to be much stronger for the PCA-based analyses. These analyses support our general hypothesis that combining single biomarkers by PCA can increase power (and perhaps specificity) in the context of genetic association analyses, but this gain in power appears to be outcome-dependent.

Lastly, we compared the results of our primary GWAS meta-analyses which were computed using fixed-effect models to analyses using random effects models (Additional file 1: Table S8). This comparison revealed that in all but one (i.e., rs12670437) analyses the statistical support of our top GWAS findings was highly comparable, suggesting that the choice of statistical model for the meta-analyses does not appreciably change our top GWAS findings.

Post-GWAS analyses on main-effect results

Our main efforts in the post-GWAS analyses concentrated on the further characterization and replication of the novel association between non-AD synaptic

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 15 of 23

functioning and markers in the *GRIN2D* region. These analyses yielded independent support of the involvement of *GRIN2D* in AD in many but not all analyses. First, our original GWAS was performed in two independent datasets (EMIF-AD and ADNI) and we note that the *GRIN2D* signal is quite consistent in both datasets (both in terms of effect size and statistical support; Additional file 1: Table S9a). Second, TWAS analyses combining large eQTL (i.e., brain from ROS/MAP) and GWAS summary statistics on AD risk and related brain phenotypes, support our conclusion of a relevant association signal in the *GRIN2D* region and suggest that the molecular effects may be mediated by affecting the expression on *GRIN2D* (see section " Functional characterization and related phenotypes", Additional file 1: Table S10). Third, while analyses in the Knight-ADRC dataset revealed no strong association signals between markers in *GRIN2D* and PC5 (Additional file 1: Table S9b), we note that three out of seven top *GRIN2D* SNPs from our primary analyses (i.e., all showing $P < 1 \times 10^{-5}$ in our GWAS) were not available in the Knight-ADRC data and could hence not be tested. Although these (and other) top markers were lacking, we were able to assess local heritability metrics in this region and observed evidence that the remaining markers explain a significant portion of PC5 variance (6.3%, $p = 0.0002$) in the Knight-ADRC data (Additional file 1: Table S4). Furthermore, inspection of all available 228 SNPs in the Knight-ADRC dataset in the general *GRIN2D* region (i.e., within the 100-kb interval from chr19: 48,365,921–48,465,818) revealed the strongest association between exons 7 and 12 of the *GRIN2D* gene, but no other gene in the region (best SNP: rs74459994; P-value $= 1.6 \times 10^{-3}$; Additional file 1: Table S9c). Furthermore, genetic effects in the Knight-ADRC individuals showed a trend in the same direction as in EMIF/ADNI with a genetic correlation of 0.42 (95% CI [− 0.46; 1.00], $p = 0.30$) between discovery and replication. While the wide confidence interval does not allow for generalized statements about the consistency of effects, the positive genetic correlation provides descriptively more context to local heritability estimates being the consequence of genetic effects in the same direction in both discovery and replication.

In summary, our extensive replication analyses provide considerable—but not unequivocal—independent support for a significant association with markers in the *GRIN2D* region and several AD-relevant phenotypes. Notwithstanding the supporting evidence and its compelling functional candidacy, the *GRIN2D* association should be considered preliminary until it is more fully characterized in future work.

Sex-specific effects

Several additional SNPs showed genome-wide significance only in the male or female subsamples (genome-wide significant sex-stratified p-value (p)), or a sex interaction effect (genome-wide significant interaction term p-value (p_{int})). For instance, rs114211800, an intronic variant of the non-coding RNA gene *LOC105377684* on chromosome 5q33.3 showed strong association with the injury/inflammation component (PC3) in males ($\beta = -1.23$SD, SE $= 0.22$, $p = 1.7 \times 10^{-8}$), but not in females ($\beta = -0.11$SD, SE $= 0.26$, $p = 0.69$, $p_{int} = 0.002$, MAF $= 0.02$). Similarly, the intergenic SNP rs12670437 (chromosome 7q11.23) was strongly associated with non-AD inflammation in male participants ($\beta = -1.20$SD, SE $= 0.22$, $p = 2.4 \times 10^{-8}$), but not in females ($\beta = +0.20$SD, SE $= 0.21$, $p = 0.33$, $p_{int} = 0.0001$, MAF $= 0.02$). Vice versa, rs140169162, located in an intron of the *MYO1D* gene on chromosome 17q11.2, showed highly specific effects on the component capturing tau pathology/degeneration (PC1) in females ($\beta = +1.61$SD, SE $= 0.30$, $p = 4.9 \times 10^{-8}$), but not in males ($\beta = -0.02$SD, SE $= 0.26$, $p = 0.94$, $p_{int} = 0.0003$, MAF $= 0.02$). Likewise, for the component tagging non-AD inflammation (PC4), rs16974493 (intergenic, chr. 13q33.3) and rs150326618 (intergenic, chr. 14q22.1) were only genome-wide significant in females (rs16974493: $\beta = -0.97$SD, SE $= 0.17$, $p = 2.2 \times 10^{-8}$; rs150326618: $\beta = +1.39$SD, SE $= 0.25$, $p = 2.6 \times 10^{-8}$), but not in males (rs16974493: $\beta = +0.18$SD, SE $= 0.22$, $p = 0.42$, $p_{int} = 0.005$, MAF $= 0.03$; rs150326618: $\beta = +0.09$SD, SE $= 0.24$, $p = 0.72$, $p_{int} = 0.001$, MAF $= 0.02$).

In addition to these sex-specific results, the sex interaction models revealed two SNPs eliciting significant evidence for sex interaction reflecting their opposite effects in males vs. females: rs1638675 (intergenic, chr. 10q25.3) and rs56194026 (approx. 2 kb upstream of *ARID3B* on chr. 15q24.1). For SNP rs1638675, the A allele showed a negative association with Aβ pathology in males ($\beta = -0.33$SD, SE $= 0.07$, $p = 1.0 \times 10^{-6}$), but positive in females ($\beta = +0.18$SD, SE $= 0.06$, $p = 4.2 \times 10^{-3}$, $p_{int} = 3.5 \times 10^{-8}$, MAF $= 0.43$). In case of rs56194026, the association was sex-dependent for the injury/inflammation component (PC3) where the C allele showed a negative effect in males ($\beta = -0.50$SD, SE $= 0.12$, $p = 1.6 \times 10^{-5}$), but a positive effect in females ($\beta = +0.41$SD, SE $= 0.12$, $p = 5.3 \times 10^{-4}$, $p_{int} = 4.7 \times 10^{-8}$, MAF $= 0.08$).

Finally, we highlight a suggestive sex difference for rs2302634 in *TMEM106B*. The effect of this SNP on the component capturing injury/inflammation (PC3) was approximately twice as large in males compared to females ($\beta_{male} = +0.34$SD vs $\beta_{female} = +0.18$SD), although

this difference did not attain statistical significance ($p = 0.06$).

Mediation analyses

Our main GWAS main analyses identified several loci showing highly significant association with the biomarker PCs defined for this study. As three of the five biomarker PCs are independently associated with diagnostic status, this raises the questions as to whether the SNP effects on PC levels also significantly impact AD development. Overall, we observed two distinct mediation patterns: (1) SNPs that affect AD either via alteration in both the Aβ pathology and tau pathology/degeneration components (*APOE*) or (2) SNPs that affect AD via the injury/inflammation PC only (*TMEM106B* and *CHI3L1*).

In the case of *APOE*, the rs429358 ε4 allele was associated with a 0.39SD higher latent AD score (SE = 0.06, $p = 4.6 \times 10^{-12}$). The mediation model suggests that this adverse effect can be partitioned into three pathways: (1) Approximately one third is attributable to mediation via the Aβ pathology component (PC2), (2) nearly one quarter of the SNP effects were due to mediation via the tau pathology/neurodegeneration component (PC1), while (3) the remaining almost 50% were due to pathways not represented by any of the measured biomarker PCs (Table 3).

The mediation analyses also suggested that increases in injury/inflammation scores (PC3) due to the T allele in SNP rs2302634 (*TMEM106B)* resulted in a significant increase of latent AD ($\beta = +0.07SD$, SE = 0.02, $p = 1.2 \times 10^{-5}$). Furthermore, we did not find evidence that *TMEM106B* affects AD by any other pathway, either measured or unmeasured. The positive mediation effect is consistent with the strong positive total AD risk effect for rs2302634 ($+Z = 5.38$, $p = 7.3 \times 10^{-8}$), and overall genome-wide significant association with AD risk recently described by two GWAS [42, 46]. The local genetic correlation between injury/inflammation scores (PC3) and AD was 0.92 (95% CI 0.50–1.00, $p = 0.001$) and 0.98 (95% CI [0.60–1.00], $p = 0.0004$) based on summary statistics from the same GWASs. See Fig. 4 for a regional plot visualizing associations between *TMEM106B* with the biomarker PC and previously reported associations with AD.

The SNP rs7551263 in the intron of *CHI3L1* was primarily associated with the component capturing non-AD inflammation (PC4). As this PC did not correlate with latent AD, we also found no evidence for mediation of AD risk via this pathway. However, rs7551263 was also nominally associated with the injury/inflammation component (PC3) (T allele: $\beta = +0.21SD$, SE = 0.06, $p = 2.3 \times 10^{-4}$) and showed evidence for mediation through this pathway ($\beta = +0.05SD$, SE = 0.02, $p = 4.0 \times 10^{-4}$). Interestingly, the

T allele was negatively associated with non-AD inflammation (PC4), so our results suggest that while the T allele decreases levels of non-AD inflammation biomarker profiles this has no measurable protective effect on latent AD. At the same time, this allele significantly increases injury/inflammation profiles, which results in a significantly higher AD risk.

Comparison to rare variant results

In contrast to our previous work based on WES-derived rare variants in a subset of the EMIF-AD dataset analyzed here [3], we found no evidence for an association between the analogous CSF biomarker components and common variants in the genes previously highlighted (i.e., *IFFO1*, *DTNB*, *NLRC3*, *SLC22A10*, *GABBR2*, and *CASZ*). Similarly, combining common SNP effects in gene-based tests did not reveal any significant associations at these loci either (Table 4). Together, these results suggest that common variants (MAF ≥ 0.01) do not appreciably contribute to the rare variant association signals identified earlier by our group.

Functional characterization and related phenotypes

The fine-mapping analyses on the lead PC5 association identified three genes through which SNPs in the *GRIN2D* locus may affect non-AD synaptic functioning (PC5). Specifically, results of the TWAS analyses suggest that *GRIN2D* SNPs lower synaptic functioning (i.e., lead to higher non-AD synaptic functioning PC5 scores) by decreasing expression of *ZSWIM9* ($Z = -5.57$, $p = 1.3 \times 10^{-8}$) and *GRIN2D* ($Z = -5.03$, $p = 1.5 \times 10^{-7}$), as well as by increasing the expression of *SLC17A7* ($Z = +4.74$, $p = 1.1 \times 10^{-6}$). The same expression profile, i.e., a predicted decrease of *GRIN2D* and *ZSWIM9* levels but increased *SLC17A7* levels, was also associated with increased risk for AD, lower educational attainment, lower cognitive ability, and higher risk of major depressive disorder across several independent studies (Additional file 1: Table S10). The statistically most robust associations were observed for decreased levels of *GRIN2D* and AD, as well as increased *ZSWIM9* / decreased *SLC17A7* levels and educational attainment (Additional file 1: Table S10).

Finally, we used the PCA-based GWAS results as input for gene-set enrichment analyses as implemented in FUMA. While these revealed a few interesting pathways and biological processes of potential relevance to AD (e.g., "regulation of macrophage chemotaxis" for PC4 [non-AD inflammation], "negative regulation of glial cell proliferation" & "neuronal differentiation" for PC3 [injury/inflammation], and "regulation of lipid transport" for PC2 [Aβ Pathology]), we note that only one ("protein localization to ciliary transition zone" [PC2]) remained

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 17 of 23

significant after multiple testing correction by FDR (Additional file 1: Table S11).

Discussion

In this work, we comprehensively explored the influence of common variants on multivariate combinations of AD CSF biomarkers representing different disease processes. In addition to confirming several previously reported GWAS loci, we identified one new region (containing *GRIN2D* and other plausible candidate genes) showing strong association with synaptic functioning in an elderly population. Furthermore, our results provide evidence for the presence of numerous loci with sex-specific effects.

Arguably the most interesting finding of our main GWAS analyses is the discovery of genome-wide significant gene-based association with variants in *GRIN2D* on chromosome 19q13.33 and the non-AD synaptic functioning component (PC5), mainly driven by Ng levels. Interestingly, SNPs in the chromosomal regions immediately flanking *GRIN2D* showed an even stronger association with non-AD synaptic functioning than variants within *GRIN2D* itself, possibly suggesting that gene expression rather than gene (dys)function may be the lead mechanism underlying this potential association, a notion that is also supported by our TWAS findings. Based on predicted gene expression by TWAS, the results suggest that the PC5-associated SNPs may lower synaptic functioning by decreasing levels of *GRIN2D*, but they may also affect other nearby genes *ZSWIM9* (decreased expression) and *SLC17A7* (increased expression).

GRIN2D encodes the GluN2D subunit of the glutamate receptor NMDAR, which plays an important role in learning and memory [51]. While mutations in *GRIN2D*-have been reported to cause epileptic encephalopathy [51], this gene's role in AD and other traits is less clear. While some recent data suggest that *GRIN2D* mRNA expression is lower in the temporal cortex of AD cases according to the AMP-AD project [52], there are no strong GWAS-based association signals reported in this region of chromosome 19 and relevant cognitive traits in the GWAS catalog (https://www.ebi.ac.uk/gwas/regions/chr19:48361628-48476971), except for an association with self-reported mathematical ability. Look-up of our two lead variants in the *GRIN2D* region (i.e., rs275844 and rs3107911) in summary statistics of two recent AD GWAS suggest a weak association for rs275844 in only one of the GWAS (*p* = 0.03) [42]. Our mediation analyses did not identify significant effects on latent AD, either, suggesting no or weak association of the locus with AD risk. In contrast, the association evidence between predicted gene expression and brain-related outcomes by TWAS was very consistent. Risk for AD, low educational

attainment, low cognitive ability and risk for major depressive disorder were all related to lower *GRIN2D* expression (and the expression of two other nearby genes).

Besides implying the *GRIN2D* locus in synaptic functioning, the current study also provides further insights into how other known AD loci may affect disease risk. This was made possible by our multivariate approach which allowed for the quantification and disentanglement of different mediator mechanisms. Specifically, in the mediation analyses, we observed two CSF biomarker profiles associated with AD, which are determined by two different gene sets. The first profile (PC1 and PC2) is characterized by decreased amyloid and increased tau, as well as increased Ng and YKL-40 levels, but not NfL. This component was most strongly associated with SNPs in the *APOE* region, in particular the well-known AD risk variant ε4. Our data suggest that the association with this variant may increase AD risk by being a catalyst for amyloid deposition or as inhibitor of amyloid clearance, represented here by the Aβ pathology PC. The resulting amyloid aggregation is thought to cascade into several neurodegenerative processes, involving formation of tau tangles, loss of synaptic functioning and inflammation [53]. Astonishingly, our results suggest that combinations of Aβ, tau, Ng, and YKL-40 assessments are able to capture most of these neurodegenerative processes triggered by the *APOE* locus, as they mediated 54–79% of the *APOE* SNP effects. However, it is important to note that the four CSF biomarkers captured by PC1 and PC2 are not sufficient to explain all genetic risk effects on AD. The second most relevant CSF biomarker pattern in that regard was the injury/inflammation component (PC3) represented by increased NfL and YKL-40 levels. PC3 levels are statistically independent of the changes in amyloid and tau levels (captured by PC1 and PC2), typically observed in AD, but associated with AD diagnostic status to a similar degree. Given prior knowledge of the non-specificity of NfL with respect to AD pathogenesis [6], we interpret this pattern to represent an independent non-AD-specific neurodegenerative pathway for dementias in general. Genetically, our results suggest that this pathway is not determined by variants in the *APOE* locus, but instead by variants in *TMEM106B* and potentially *CHI3L1*. *TMEM106B* affects neuronal loss [54] and has been convincingly associated with risk for at least two forms of dementias, i.e., fronto-temporal dementia (FTD) [54] and AD [13, 42, 55]. Here, the lead variant for *TMEM106B* is rs2302634, which is in perfect LD ($r^2=1$) with the lead variant from our recently published NfL-specific GWAS in the same two datasets (i.e., rs7797705) [13]. Both variants are in either perfect ($r^2=1$) or near-perfect ($r^2=0.98$) LD with the lead variant reported to be

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 18 of 23

a risk factor for FTD, i.e., rs1990622, suggesting that the signal underlying PC3 (this study) and risk for FTD (previous work) are strongly correlated and may be elicited by the same underlying causal variant. *CHI3L1* encodes the YKL-40 protein and we and others have previously demonstrated that common and possibly rare variants at this locus represent *cis*pQTLs of CSF YKL-40 [3, 13]. The novel results from our current study suggest that *CHI3L1* variants may be associated with increased neuronal injury and inflammation leading to a heightened AD risk. We note, however, that the associations between *CHI3L1* and injury/inflammation (PC3) showed opposite effect directions from those between *CHI3L1* and non-AD inflammation (PC4). Comparisons with univariate analyses of YKL-40 suggest, that the single biomarker analyses do not allow to distinguish between clinically relevant YKL-40 levels co-occurring with NfL and non-clinically relevant variation, and thus may mask these complex relationships. This may result in underestimated effect sizes and emphasizes the advantages of applying a multivariate approach. In addition, it is important to emphasize that the association signals elicited by the *CHI3L1* variant for the mediation effect of PC3 on AD risk (i.e., rs7551263; Table 3) are not correlated ($r^2 = 0.04$) with the lead *CHI3L1* variants of the PC4 main effect (e.g., rs10399931; Table 2). Interestingly, both variants were highlighted as pQTL of YKL-40 in previous work from our group [13].

Additional insights resulted from the sex-stratified analyses which revealed several SNPs either showing associations in only one sex stratum, or opposite effects in males and females. As examples, we highlight two such SNPs: First, rs140169162 is located in an intron of *MYO1D* and showed strong association with tau pathology/neurodegeneration (PC1) with evidence for a mediation effect on latent AD, but only in the female subsample. Interestingly, SNPs in *MYO1D* has been found to have a female-specific effect on hernias, as well [56]. Despite its apparent sex-specificity previous work has nominated *MYOD1* as a potential drug target for AD according to predictive network analysis [57, 58]. The encoded protein, myogenic differentiation 1, is involved in myelin sheath formation [59] and both common [60] and rare [61] variants have been associated with autism, supporting *MYO1D*'s role in neural development and functioning.

Second, rs56194026 is located near *ARID3B* and was associated with injury/inflammation (PC3), with strong but opposite effects in males vs. females. The gene encodes AT-rich interaction domain 3B, a DNA-binding protein from the ARID family of proteins which are involved in embryonic patterning, cell lineage gene regulation, cell cycle control, transcriptional regulation,

and possibly in chromatin structure modification [62]. Samyesudhas et al. [63] recently suggested a relevant role of this protein in AD development, as *ARID3B* is expressed in response to the amyloid precursor protein intracellular domain and neuronal injury. However, it remains unclear why SNPs near this gene would have opposite effects in males and females. Possibly, this is related to the higher mean NfL levels in males, or the genes' proposed function as regulator of sex-biased expression [64].

A major strength of our study is the application of multivariate analyses based on five CSF biomarker profiles and the estimation of mediation effects. Studying component patterns of different biomarker combinations allows to shed new light and provide new insights on how common genetic variants affect biomarkers and AD risk beyond their effects on the levels of single biomarkers. In the context of our study, PCA-based GWAS analyses showed stronger associations when compared to single biomarkers for all PC phenotypes except PC2. Further support for the multivariate approach stems from our results with YKL-40 and Ng which show different association patterns, depending on whether or not they co-vary with the levels of other biomarkers. The inclusion of the X-chromosome and examination of sex differences are additional strengths of our study. While no SNPs on the X-chromosome attained genome-wide significance, we identified several SNPs showing sex-specific effects. This highlights the importance of modeling sex interactions, especially for biomarkers with pronounced differences in mean levels.

In addition to these strengths, we note the following potential limitations. First, while our sample size is generally large for a CSF biomarker study, it is small compared to GWAS of other complex traits, including recent GWAS in the AD field [42, 46]. Second, the sample size limitation is aggravated in the sex-specific analyses and SNPs with low MAF, which need to be interpreted with caution and require further replication. Third, we only studied individuals of European descent. It remains unclear whether and to which degree our results are relevant also in non-European ancestries. It is well-known, for instance, that the *APOE* risk effects on AD are ancestry-dependent [65, 66]. Fourth, it is important to emphasize, that our mediation analyses are based on the assumption that the analyzed CSF biomarkers reflect pathological processes that *precede* and *cause* AD symptoms. An alternative—and often times equally plausible—interpretation is that the uncovered SNPs affect AD symptoms independently of biomarker levels and that the component associations observed here actually reflect a *consequence* of AD pathogenesis. While the specificity of the mediation results using certain PCs but

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 19 of 23

not others generally supports the assumed causal directions, longitudinal studies, e.g., on MCI conversion, are needed to confirm the findings of this arm of our project. Fifth, we note that one of our lead signals, i.e., the association between the non-AD synaptic functioning PC (mainly driven by Ng levels) and variants in *GRIN2D*, only showed strong evidence for association in two out of three CSF biomarker datasets. While additional analyses (e.g., TWAS using several AD-relevant phenotypes, local heritability estimates) generally support the *GRIN2D* association, this signal should be considered preliminary until assessed in additional datasets of sufficient size. Sixth, our primary GWAS analyses were performed with five independent CSF biomarker phenotypes. In agreement with common practice in the GWAS field [67–69], we did not adjust for testing multiple traits in our study, but define genome-wide significance at the conventional threshold of 5×10^{-8}. Accounting for all five traits would theoretically lead to a study-wide alpha level of 1×10^{-8}, which is surpassed by all identified GWAS main effects except those at *GRIN2D* (for which the smallest SNP-based P-value is 5×10^{-7}). Thus, the decision to refrain from adjusting for trait multiplicity does not change any of the main conclusions of our study.

Conclusions

In summary, in this first multivariate CSF biomarker GWAS, we observed at least one novel locus showing strong and convincing association with non-AD specific biomarker patterns. The results also suggest the presence of two distinct mediation pathways, by which common SNPs may affect AD risk. One pathway is related to amyloid and tau pathology and is mostly determined by *APOE* SNPs. The second pathway is related to increased neuronal injury and inflammation, captured by NfL and YKL-40. Genetically, this latter pathway is mostly driven by variants in *TMEM106B* and *CHI3L1*. Pathway-aware genetic studies with larger sample sizes and in more diverse ancestries are needed to further understand the complex etiology of AD and to translate genetic information to personalized medicine approaches.

Abbreviations

AD	Alzheimer's disease
CSF	Cerebrospinal fluid
PCA	Principal component analysis
EMIF-AD MBD	European Medical Information Framework for Alzheimer's Disease Multimodal Biomarker Discovery
ADNI	Alzheimer's Disease Neuroimaging Initiative
pTau	Phosphorylated tau
Ng	Neurogranin
NfL	Neurofilament light chain
MCI	Mild cognitive impairment
SEM	Structural equation model
WLSMV	Weighted least square mean and variance adjusted

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-023-01233-z.

Additional file 1: Table S1. Participant Characteristics CSF Biomarker Sample. **Table S2.** Participant Characteristics GWAS Discovery Sample. **Table S3.** Participant Characteristics GWAS Replication Sample. **Table S4.** Local heritability and variance explained. **Table S5.** PCA loadings (individual cohorts). **Table S6.** Sex differences in CSF biomarker PCs and associations with Alzheimer's disease. **Table S7.** Comparison of PCA-based GWAS results vs. single biomarker GWAS results. **Table S8.** Comparison of meta-analysis results using fixed effect vs. random effects models. **Table S9a.** PC5 top hits in meta-analysis split by discovery datasets. **Table S9b.** PC5 non-AD synaptic functioning replication and meta-analysis for top SNPs ($p < 10E-3$). **Table S9c.** SNP-based association results in Knight-ADRC dataset in GRIN2D region (bold font highlights nominally significant signals). **Table S10.** TWAS Fine-mapping analyses. **Table S11.** Gene-set enrichment analyses based on PC-based GWAS (top 20 gene sets [by *P*-value] are shown; bold font = significant after FDR [0.05]).

Additional file 2: Supplementary Methods and Figures. Additional methods on genotyping, multiple testing adjustment, replication study and figures S1-S5.

Authors' contributions

AN, OO, KS, and LB conceptualized the study and drafted the manuscript. LB and KS supervised the study. OO performed statistical analyses (GWAS, meta-analysis, follow-up). AN performed statistical analyses (PCA, SEM, TWAS fine-mapping, follow-up). JT performed replication analyses supervised by CC. FK assisted with and advised on statistical analyses. Other authors contributed to data collection and/or clinical and biomarker measurements and data cleaning. All authors read and approved the final manuscript.

Availability of data and materials
In accordance with EU law and participant privacy, clinical individual-level data from EMIF-AD is not available publicly, but can be obtained via EMIF-AD (https://emif-catalogue.eu; http://www.emif.eu/about/emif-ad) [70]. Registered users can download ADNI data from http://adni.loni.usc.edu/ [71]. The results published here are in part based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.org) [72]. ROS/MAP predicted gene expression weights can be obtained from https://doi.org/10.5281/zenodo.5745927 [73]. Analysis code can be found at https://github.com/aneumann-science/common_variants_csf_biomarkers [74]. Summary statistics for all GWAS analyses are available for download at URL: https://doi.org/10.5281/zenodo.8334941 [75].

Declarations

Ethics approval and consent to participate
Written informed consent was obtained for all assessment before the start of the study [4]. The study was conducted in accordance to the Declaration of Helsinki and ethical approval was obtained from the Ethical Committee of the University of Lübeck, as well as local committees of consortium members [4]. More details on the recruitment and phenotype ascertainment protocols used in the EMIF-AD dataset can be found in Bos et al. [4].

Consent for publication
Not applicable.

Competing interests
FB is on the steering committee or iDMC member for Biogen, Merck, Roche, EISAI, and Prothena. consultant for Roche, Biogen, Merck, IXICO, Jansen, Combinostics, has research agreements with Merck, Biogen, GE Healthcare, Roche, and is co-founder and shareholder of Queen Square Analytics LTD; HZ has served at scientific advisory boards and/or as a consultant for Abbvie, Alector, ALZPath, Annexon, Apellis, Artery Therapeutics, AZTherapies, CogRx, Denali, Eisai, Nervgen, Novo Nordisk, Passage Bio, Pinteon Therapeutics, Red Abbey Labs, reMYND, 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 (outside submitted work). AL is on the editorial board of Neurology and Brain Communications, has served at scientific advisory boards from Fujirebio-Europe, Nutricia, Roche-Genentech, Biogen, Grifols, and Roche Diagnostics, and has filed a patent application of synaptic markers in neurodegenerative diseases. D.A. participated in advisory boards from Fujirebio-Europe and Roche Diagnostics and received speaker honoraria from Fujirebio-Europe, Roche Diagnostics, Nutricia, Krka Farmacéutica S.L., Zambon S.A.U., and Esteve Pharmaceuticals S.A. D.A. declares a filed patent application (WO2019175379 A1 Markers of synaptopathy in neurodegenerative disease). SL is currently an employee of Janssen Medical Ltd (UK), a co-founder of Akrivia Health Ltd (UK) and within the past 5 years has filed patents related to biomarkers unrelated to the current work and advised or given lectures for Merck, Optum Labs, and Eisai as well as having received grant funding from multiple companies as part of EU IMI programs and from Astra Zeneca. JP received consultation honoraria from Nestle Institute of Health Sciences, Ono Pharma, OM Pharma, and Fujirebio, unrelated to the submitted work. SE has served on scientific advisory boards for Biogen, Danone, icometrix, Novartis, Nutricia, and Roche and received unrestricted research grants from Janssen Pharmaceutica and ADx Neurosciences (paid to institution). JR was an employee at GSK and currently an employee at the MSD London Discovery Centre, U.K. CC has received research support from: GSK and EISAI. CC is a member of the advisory board of Vivid Genomics and Circular Genomics and owns stocks. The remaining authors declare that they have no competing interests.

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 21 of 23

Author details

[1]Complex Genetics of Alzheimer's Disease Group, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium. [2]Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium. [3]Child and Adolescent Psychiatry/Psychology, Erasmus University Medical Center, Rotterdam, Netherlands. [4]Lübeck Interdisciplinary Platform for Genome Analytics (LIGA), University of Lübeck, Ratzeburger Allee 160, V50.2M, Lübeck 23562, Germany. [5]Netherlands Institute for Health Services Research, Utrecht, Netherlands. [6]Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA. [7]NeuroGenomics and Informatics Center, Washington University School of Medicine, St Louis, MO, USA. [8]Alzheimer Centrum Limburg, Maastricht University, Maastricht, Netherlands. [9]Genetics and Aging Unit and McCance Center for Brain Health, Department of Neurology, Massachusetts General Hospital, Boston, MA, USA. [10]Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, Netherlands. [11]Department of Psychiatry and Neurochemistry, University of Gothenburg, Gothenburg, Sweden. [12]Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden. [13]Laboratory for Cognitive Neurology, Department of Neurosciences, KU Leuven, Leuven, Belgium. [14]Neurology Service, University Hospital Leuven, Leuven, Belgium. [15]Neurochemistry Laboratory, Department of Clinical Chemistry, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, Netherlands. [16]Department of Neurology and Memory Clinic, Universitair Ziekenhuis Brussel (UZ Brussel) and Center for Neurosciences (C4N), Vrije Universiteit Brussel (VUB), Brussels, Belgium. [17]Memory Center, Department of Rehabilitation and Geriatrics, Geneva University and University Hospitals, Geneva, Switzerland. [18]Clinical Pharmacology & Pharmacovigilance Department, Marseille University Hospital, Marseille, France. [19]Neurosciences Therapeutic Area, GlaxoSmithKline R&D, Stevenage, UK. [20]Neuroscience & Cognition, CHU de Lille, University of Lille, Inserm, France. [21]Memory Unit, Neurology Department, Hospital de Sant Pau, Barcelona, Spain. [22]Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain. [23]Department of Psychiatry, Psychotherapy and Psychosomatics, University of Zürich, Zurich, Switzerland. [24]Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland. [25]Division of Biology & Biomedical Sciences, Washington University in St. Louis, St Louis, MO, USA. [26]Center for Research and Advanced Therapies, CITA—Alzheimer Foundation, San Sebastian, Spain. [27]Zumarraga Hospital, Osakidetza, Integrated Health Organization (OSI) Goierri-Urola Garia, Basque Country, Spain. [28]Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, Boston, UK. [29]NIHR BioResource Centre Maudsley, NIHR Maudsley Biomedical Research Centre (BRC) at South London and Maudsley NHS Foundation Trust (SLaM) & Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, UK. [30]Health Data Research UK London, University College London, London, UK. [31]Institute of Health Informatics, University College London, London, UK. [32]The National Institute for Health Research University College London Hospitals Biomedical Research Centre, University College London, London, UK. [33]Steno Diabetes Center, Copenhagen, Denmark. [34]Institute of Pharmaceutical Science, King's College London, London, UK. [35]Neurodegenerative Brain Diseases Group, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium. [36]Alzheimer Center and Department of Neurology, VU University Medical Center, Amsterdam, Netherlands. [37]Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, Netherlands. [38]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [39]Ageing Epidemiology Research Unit, School of Public Health, Imperial College, London, UK. [40]Department of Radiology and Nuclear Medicine, Amsterdam UMC, Vrije Universiteit, Amsterdam, Netherlands. [41]Queen Square Institute of Neurology and Centre for Medical Image Computing, University College London, London, UK. [42]Hope Center for Neurological Disorders, Washington University School of Medicine, St Louis, MO, USA. [43]Janssen Medical Ltd, Wycombe, UK. [44]Department of Psychiatry, University of Oxford, Oxford, UK. [45]AC Immune SA, Lausanne, Switzerland. [46]Janssen R&D, LLC, Beerse, Belgium. [47]Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK. [48]UK Dementia Research Institute, University College London, London, UK. [49]Hong Kong Center for Neurodegenerative Diseases, Hong Kong, China. [50]Centre for Lifespan Changes in Brain and Cognition, University of Oslo, Oslo, Norway.

References

1. Sims R, Hill M, Williams J. The multiplex model of the genetics of Alzheimer's disease. Nat Neurosci. 2020;23:311–22.
2. Olsson B, Lautner R, Andreasson U, Öhrfelt A, Portelius E, Bjerke M, Hölttä M, Rosén C, Olsson C, Strobel G, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. Lancet Neurol. 2016;15:673–84.
3. Neumann A, Küçükali F, Bos I, Vos SJB, Engelborghs S, De Pooter T, Joris G, De Rijk P, De Roeck E, Tsolaki M, et al. Rare variants in IFFO1, DTNB, NLRC3 and SLC22A10 associate with Alzheimer's disease CSF profile of neuronal injury and inflammation. Mol Psychiatry. 2022;26:1–10.
4. Bos I, Vos S, Vandenberghe R, Scheltens P, Engelborghs S, Frisoni G, Molinuevo JL, Wallin A, Lleó A, Popp J, et al. The EMIF-AD multimodal biomarker discovery study: design, methods and cohort characteristics. Alzheimer's Res Ther. 2018;10:1–9.
5. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol. 2009;65:403–13.
6. Blennow K, Zetterberg H. Biomarkers for Alzheimer's disease: current status and prospects for the future. J Intern Med. 2018;284:643–63.
7. Mavroudis IA, Petridis F, Chatzikonstantinou S, Kazis D. A meta-analysis on CSF neurogranin levels for the diagnosis of Alzheimer's disease and mild cognitive impairment. Aging Clin Exp Res. 2020;32:1639–46.
8. Llorens F, Thüne K, Tahir W, Kanata E, Diaz-Lucena D, Xanthopoulos K, Kovatsi E, Pleschka C, Garcia-Esparcia P, Schmitz M, et al. YKL-40 in the brain and cerebrospinal fluid of neurodegenerative dementias. Mol Neurodegener. 2017;12:1–21.
9. Connolly K, Lehoux M, O'Rourke R, Assetta B, Erdemir GA, Elias JA, Lee CG, Huang YWA. Potential role of chitinase-3-like protein 1 (CHI3L1/YKL-40) in neurodegeneration and Alzheimer's disease. Alzheimer's Dement. 2023;19(1):9–24.
10. Querol-Vilaseca M, Colom-Cadena M, Pegueroles J, San Martín-Paniello C, Clarimon J, Belbin O, Fortea J, Lleó A. YKL-40 (Chitinase 3-like I) is expressed in a subset of astrocytes in Alzheimer's disease and other tauopathies. J Neuroinflammation. 2017;14:1–10.
11. Prokopenko D, Lee S, Hecker J, Mullin K, Morgan S, Katsumata Y, (ADNI), ADNI, Weiner MW, Fardo DW, Laird N, et al. Region-based analysis of rare genomic variants in whole-genome sequencing datasets reveal two novel Alzheimer's disease-associated genes: DTNB and DLG2. MedRxiv. 2021;2021.06.09.21258576.
12. Deming Y, Li Z, Kapoor M, Harari O, Del-Aguila JL, Black K, Carrell D, Cai Y, Fernandez MV, Budde J, et al. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. Acta Neuropathol. 2017;133:839–56.
13. Hong S, Dobricic V, Ohlei O, Bos I, Vos SJB, Prokopenko D, Tijms BM, Andreasson U, Blennow K, Vandenberghe R, et al. TMEM106B and CPOX are genetic determinants of cerebrospinal fluid Alzheimer's disease biomarker levels. Alzheimer's Dement. 2021;17(10):1628–40.
14. Dai QH, Gong DK. Association of the polymorphisms and plasma level of CHI3L1 with Alzheimer's disease in the Chinese Han population: a case-control study. Neuropsychobiology. 2019;77:29–37.
15. Hong S, Prokopenko D, Dobricic V, Kilpert F, Bos I, Vos SJB, Tijms BM, Andreasson U, Blennow K, Vandenberghe R, et al. Genome-wide association study of Alzheimer's disease CSF biomarkers in the EMIF-AD multimodal biomarker discovery dataset. Transl Psychiatry. 2020;10(1):403.
16. Aschard H, Vilhjálmsson BJ, Greliche N, Morange PE, Trégouët DA, Kraft P. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. Am J Hum Genet. 2014;94:662–76.
17. Galesloot TE, van Steen K, Kiemeney LALM, Janss LL, Vermeulen SH. A comparison of multivariate genome-wide association methods. PLoS One. 2014;9:e95923.
18. Toro CA, Zhang L, Cao J, Cai D. Sex differences in Alzheimer's disease: understanding the molecular impact. Brain Res. 2019;1719:194–207.
19. Mielke MM. Consideration of sex differences in the measurement and interpretation of Alzheimer disease-related biofluid-based biomarkers. J Appl Lab Med. 2020;5:158–69.
20. Dumitrescu L, Barnes LL, Thambisetty M, Beecham G, Kunkle B, Bush WS, Gifford KA, Chibnik LB, Mukherjee S, de Jager PL, et al. Sex differences in the genetic predictors of Alzheimer's pathology. Brain. 2019;142:2581.

Neumann *et al. Genome Medicine*   (2023) 15:79

Page 22 of 23

21. Xu J, Green R, Kim M, Lord J, Ebshiana A, Westwood S, Baird AL, Nevado-Holgado AJ, Shi L, Hye A, et al. Sex-specific metabolic pathways were associated with Alzheimer's disease (AD) endophenotypes in the European medical information framework for AD multimodal biomarker discovery cohort. Biomedicines. 2021;9:1610.

22. Homann J, Osburg T, Ohlei O, Dobricic V, Deecke L, Bos I, Vandenberghe R, Gabel S, Scheltens P, Teunissen CE, et al. Genome-wide association study of Alzheimer's disease brain imaging biomarkers and neuropsychological phenotypes in the European medical information framework for Alzheimer's disease multimodal biomarker discovery dataset. Front Aging Neurosci. 2022;14:195.

23. Bos I, Vos S, Verhey F, Scheltens P, Teunissen C, Engelborghs S, Sleegers K, Frisoni G, Blin O, Richardson JC, et al. Cerebrospinal fluid biomarkers of neurodegeneration, synaptic integrity, and astroglial activation across the clinical Alzheimer's disease spectrum. Alzheimer's Dement. 2019;15:644–54.

24. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48:1279–83.

25. Blennow K, Shaw LM, Stomrud E, Mattsson N, Toledo JB, Buck K, Wahl S, Eichenlaub U, Lifke V, Simon M, et al. Predicting clinical decline and conversion to Alzheimer's disease or dementia using novel Elecsys Aβ(1–42), pTau and tTau CSF immunoassays. Sci Rep. 2019;9(1):19024.

26. Mielke MM, Syrjanen JA, Blennow K, Zetterberg H, Vemuri P, Skoog I, Machulda MM, Kremers WK, Knopman DS, Jack C, et al. Plasma and CSF neurofilament light: relation to longitudinal neuroimaging and cognitive measures. Neurology. 2019;93:E252–60.

27. Wang L. Association of cerebrospinal fluid Neurogranin with Alzheimer's disease. Aging Clin Exp Res. 2019;31:185–91.

28. Spellman DS, Wildsmith KR, Honigberg LA, Tuefferd M, Baker D, Raghavan N, Nairn AC, Croteau P, Schirm M, Allard R, et al. Development and evaluation of a multiplexed mass spectrometry based assay for measuring candidate peptide biomarkers in Alzheimer's Disease Neuroimaging Initiative (ADNI) CSF. Proteomics - Clin Appl. 2015;9:715–31.

29. Auer PL, Reiner AP, Leal SM. The effect of phenotypic outliers and non-normality on rare-variant association testing. Eur J Hum Genet. 2016;24:1188–94.

30. Josse J, Husson F. missMDA: a package for handling missing values in multivariate data analysis. J Stat Softw. 2016;70:1–31.

31. Revelle W. psych: procedures for psychological, psychometric, and personality research. Evanston: Northwest. Univ.; 2014. p. 165.

32. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2023. https://www.R-project.org/.

33. Beaujean A. Latent variable modeling using R: a step-by-step guide. New York: Routledge; 2014.

34. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Softw. 2012;48:1–36.

35. Chang C, Chow C, Tellier L, Vattikuti S, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4(1):s13742–015.

36. Smith SM, Douaud G, Chen W, Hanayik T, Alfaro-Almagro F, Sharp K, Elliott LT. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. Nat Neurosci. 2021;24(5):737–45.

37. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26:2190–1.

38. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11:1–19.

39. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826.

40. Werme J, van der Sluis S, Posthuma D, de Leeuw CA. An integrated framework for local genetic correlation analysis. Nat Genet. 2022;54:274–82.

41. Essers E, Binter AC, Neumann A, White T, Alemany S, Guxens M. Air pollution exposure during pregnancy and childhood, APOE ε4 status and Alzheimer polygenic risk score, and brain structural morphology in preadolescents. Environ Res. 2023;216(Pt 2):114595.

42. Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, Rongve A, Børte S, Winsvold BS, Drange OK, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. Nat Genet. 2021;53(9):1276–82.

43. Mancuso N, Freund MK, Johnson R, Shi H, Kichaev G, Gusev A, Pasaniuc B. Probabilistic fine-mapping of transcriptome-wide association studies. Nat Genet. 2019;51:675–82.

44. De Jager PL, Ma Y, McCabe C, Xu J, Vardarajan BN, Felsky D, Klein HU, White CC, Peters MA, Lodgson B, et al. Data descriptor: a multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. Sci Data. 2018;5:1–13.

45. Allen M, Carrasquillo MM, Funk C, Heavner BD, Zou F, Younkin CS, Burgess JD, Chai HS, Crook J, Eddy JA, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. Sci Data. 2016;3:1–10.

46. Bellenguez C, Küçükali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, Naj AC, Campos-Martin R, Grenier-Boley B, Andrade V, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet. 2022;54(4):412–36.

47. Okbay A, Wu Y, Wang N, Jayashankar H, Bennett M, Nehzati SM, Sidorenko J, Kweon H, Goldman G, Gjorgjieva T, et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. Nat Genet. 2022;54:437–49.

48. Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD, Hagenaars SP, Ritchie SJ, Marioni RE, Fawns-Ritchie C, et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun. 2018;9:2098.

49. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, Coleman JRI, Hagenaars SP, Ward J, Wigmore EM, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci. 2019;22:343–52.

50. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101:5–22.

51. Camp CR, Yuan H. GRIN2D/GluN2D NMDA receptor: unique features and its contribution to pediatric developmental and epileptic encephalopathy. Eur J Paediatr Neurol. 2020;24:89–99.

52. AGORA GRIN2D. https://agora.adknowledgeportal.org/genes/(genes-router:gene-details/ENSG00000105464). Accessed 9 Aug 2022.

53. Potter H, Wisniewski T. Apolipoprotein E: essential catalyst of the Alzheimer amyloid cascade. Int J Alzheimers Dis. 2012;2012:489428.

54. Li Z, Farias FHG, Dube U, Del-Aguila JL, Mihindukulasuriya KA, Fernandez MV, Ibanez L, Budde JP, Wang F, Lake AM, et al. The TMEM106B FTLD-protective variant, rs1990621, is also associated with increased neuronal proportion. Acta Neuropathol. 2020;139:45–61.

55. Lill CM, Bertram L. Genome-wide analysis furthers decoding of Alzheimer disease genetics. Nat Rev Neurol. 2022;18(7):387–8.

56. Choquet H, Li W, Yin J, Bradley R, Hoffmann TJ, Nandakumar P, Team, 23 and Me Research, Mostaedi R, Tian C, Ahituv N, et al. Ancestry- and sex-specific effects underlying inguinal hernia susceptibility identified in a multiethnic genome-wide association study meta-analysis. Hum Mol Genet. 2022;00:1–15.

57. AGORA MYO1D. https://agora.adknowledgeportal.org/genes/(genes-router:gene-details/ENSG00000176658). Accessed 9 Aug 2022.

58. Patel KR, Zhu K, Henrion MYR, Beckmann ND, Moein S, Alamprese ML, Allen M, Wang X, Chan G, Pertel T, et al. Single cell-type integrative network modeling identified novel microglial-specific targets for the phagosome in Alzheimer's disease. BioRxiv. 2020;2020.06.09.143529.

59. Yamazaki R, Ishibashi T, Baba H, Yamaguchi Y. Knockdown of unconventional myosin ID expression induced morphological change in oligodendrocytes. ASN Neuro. 2016;8(5):1759091416669609.

60. Stone JL, Merriman B, Cantor RM, Geschwind DH, Nelson SF. High density SNP association study of a major autism linkage region on chromosome 17. Hum Mol Genet. 2007;16:704–15.

61. Ou J, Liu R, Shen Y, Xia K, Zhao J. An overview of genetic and environmental risk of autism spectrum disorder. Glob Clin Transl Res. 2019;1:37–44.

62. Homo sapiens AT-rich interaction domain 3B (ARID3B), transcript varian - Nucleotide - NCBI. https://www.ncbi.nlm.nih.gov/nuccore/1677496623. Accessed 9 Aug 2022.

63. Samyesudhas SJ, Roy L, Cowden Dahl KD. Differential expression of ARID3B in normal adult tissue and carcinomas. Gene. 2014;543:174–80.

64. Deegan DF, Karbalaei R, Madzo J, Kulathinal RJ, Engel N. The developmental origins of sex-biased expression in cardiac development. Biol Sex Differ. 2019;10:1–20.

65. Naslavsky MS, Suemoto CK, Abreu Brito L, Oliveira Scliar M, Ferretti-Rebustini Rn RE, Diehl 1 0 Rodriguez R, Leite REP, Matta NA, Borda V, et al. Global and local ancestry modulate APOE association with Alzheimer's neuropathology and cognitive outcomes in an admixed sample. MedRxiv. 2022;2022.02.02.22270331.

66. Griswold AJ, Celis K, Bussies PL, Rajabli F, Whitehead PL, Hamilton-Nelson KL, Beecham GW, Dykxhoorn DM, Nuytemans K, Wang L, et al. Increased APOE ε4 expression is associated with the difference in Alzheimer's disease risk from diverse ancestral backgrounds. Alzheimer's Dement. 2021;17:1179–88.

67. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. Nat Genet. 2018;50:1593–9.

68. Althouse AD. Adjust for multiple comparisons? It's not that simple. Ann Thorac Surg. 2016;101:1644–5.

69. Feise RJ. Do multiple outcome measures require p-value adjustment? BMC Med Res Methodol. 2002;2:1–4.

70. EMIF. EMIF catalogue. 2019. https://emif-catalogue.eu.

71. Alzheimer's Disease Neuroimaging Initiative. ADNI data. 2017. http://adni.loni.usc.edu/.

72. Accelerating Medicines Partnership in Alzheimer's Disease (AMP-AD) program. The RNAseq harmonization study (RNAseq harmonization). AD Knowledge Portal; 2022. https://adknowledgeportal.synapse.org/Explore/Studies/DetailsPage/StudyDetails?Study=syn21241740.

73. Küçükali F, Sleegers K. Significant eQTLs and expression TWAS reference panels (AMP-AD brain and EADB Belgian LCL cohorts). ZENODO; 2021. https://doi.org/10.5281/zenodo.5745927.

74. Neumann A. Analysis scripts for "Multivariate GWAS of Alzheimer's disease CSF biomarker profiles implies GRIN2D in synaptic functioning". GitHub; 2023. https://github.com/aneumann-science/common_variants_csf_biomarkers.

75. Neumann A, Ohlei O, Sleegers K, Bertram L. Summary statistics. ZENODO; 2023. https://doi.org/10.5281/zenodo.8334941.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.