Explainable artificial intelligence identifies an AQP4 polymorphism-based risk score associated with brain amyloid burden

Simone Beer [a,*], David Elmenhorst [a,b], Gerard N. Bischof [a,b], Alfredo Ramirez [c,d,e,f,g], Andreas Bauer [a], Alexander Drzezga [a,b,d], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Institute of Neuroscience and Medicine (INM-2), Forschungszentrum Jülich, Germany
[b] Department of Nuclear Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany
[c] Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany
[d] German Center for Neurodegenerative Diseases (DZNE), Bonn-Cologne, Germany
[e] Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany
[f] Department for Neurodegenerative Diseases and Geriatric Psychiatry, Bonn, Germany
[g] Department of Psychiatry and Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, United States

ARTICLE INFO

ABSTRACT

Aquaporin-4 (AQP4) is hypothesized to be a component of the glymphatic system, a pathway for removing brain interstitial solutes like amyloid-β (Aβ). Evidence exists that genetic variation of AQP4 impacts Aβ clearance, clinical outcome in Alzheimer's disease as well as sleep measures. We examined whether a risk score calculated from several AQP4 single-nucleotide polymorphisms (SNPs) is related to Aβ neuropathology in older cognitively unimpaired white individuals. We used a machine learning approach and explainable artificial intelligence to extract information on synergistic effects of AQP4 SNPs on brain amyloid burden from the ADNI cohort. From this information, we formulated a sex-specific AQP4 SNP-based risk score and evaluated it using data from the screening process of the A4 study. We found in both cohorts significant associations of the risk score with brain amyloid burden. The results support the hypothesis of an involvement of the glymphatic system, and particularly AQP4, in brain amyloid aggregation pathology. They suggest also that different AQP4 SNPs exert a synergistic effect on the build-up of brain amyloid burden.

1. Introduction

The water channel aquaporin 4 (AQP4) is a protein which is essential for the regulation of water homeostasis in the brain (Salman et al., 2022). AQP4 is highly expressed in astrocytes, mainly polarized on their perivascular endfeet. It is considered to play a crucial role in the glymphatic system, a brain-wide perivascular network which is believed to facilitate the clearance of waste products like amyloid-β (Aβ) from the brain (Iliff et al., 2012; Valenza et al., 2020). Accumulation of Aβ, followed by neuritic plaque formation, is one of the neuropathological hallmarks of Alzheimer's disease (AD). An impaired ability to clear Aβ from the brain via the glymphatic system, the brain's analogue to the

lymphatic system of the body, could contribute to its accumulation (Nedergaard and Goldman, 2020). The glymphatic system is mostly active during sleep, and it was shown in mice that glymphatic influx and clearance exhibit endogenous, circadian rhythms, supported at least in part via rhythmic localization of AQP4 to the perivascular endfeet of astrocytes (Hablitz et al., 2020).

In a human postmortem study it was shown that AQP4 expression increased in the aging brain, and a loss of perivascular localization was significantly associated with AD status and increased Aβ burden. However, perivascular AQP4 localization was preserved among the eldest individuals who remained cognitively intact (Simon et al., 2022; Zeppenfeld et al., 2017), suggesting that a loss of perivascular AQP4

localization might contribute to the development of AD pathology. Besides glymphatic fluid clearance, increasing evidence suggests that AQP4 is also involved in brain inflammation, synaptic plasticity and memory formation, as well as regulation of extracellular space volume and potassium homeostasis (Hubbard et al., 2018; Mader and Brimberg, 2019).

AQP4 is encoded by the *AQP4* gene. Genetic variation of *AQP4* may impact the expression and functionality of the water channels and therefore the clearance of Aβ. For example, Woo et al. (2018) examined the functionality of one *AQP4* single-nucleotide polymorphism (SNP) in modulating the expression level of *AQP4* in an *in vitro* luciferase reporter assay, and found this SNP to cause *AQP4* expression-level changes.

Recent evidence indicates that genetic variation of *AQP4* is associated with measures of dementia. Burfeind et al. (2017) have studied the effects of five *AQP4* SNPs in cognition and disease progression of AD, and have found associations with altered rates of cognitive decline after AD diagnosis. Two SNPs were associated with slower cognitive decline, two with more rapid cognitive decline, and one was associated with slower cognitive but faster functional decline. The authors also report a possible synergistic effect of the *AQP4* SNPs. Chandra et al. (2021) have examined eighteen *AQP4* SNPs, and found one which was associated with decreased brain amyloid load, measured as uptake of the amyloid tracer [^{18}F]Florbetapir, as well as one with increased amyloid load, disease stage progression and cognitive decline in Aβ-positive late mild cognitive impairment (MCI) and AD patients. Fang et al. (2022) have studied the clinical implication of eleven *AQP4* SNPs in Parkinson's disease (PD). They have found one SNP which was associated with slower dementia conversion, better performance in cognitive tests and lower Aβ deposition, and one SNP with faster progression to MCI and worse performance in cognitive tests. One of the SNPs also modulated an association of REM sleep behaviour disorder and the biomarker CSF Aβ42. Overall, these findings suggests that AQP4 indeed modulates cognitive disease trajectories in AD and MCI patients, potentially through moderating the build-up of disease pathology.

Recent work also describes a link between AD, sleep and AQP4. Rainey-Smith et al. (2018) report an association of *AQP4* SNPs with self-reported sleep quality and a moderation of the relationship between sleep latency, duration, and brain Aβ burden. Ulv Larsen et al. (2020) showed that a haplotype of *AQP4* SNPs, containing one of the SNPs which was associated with cognition and function in Burfeind et al. (2017), is also associated with slow wave energy, a combined measure of sleep intensity and duration. Despite these findings, even the latest genome-wide association studies (GWAS) did not identify *AQP4* as a risk locus for Alzheimer's disease (Bellenguez et al., 2022), suggesting a potential involvement of *AQP4* variants rather in endophenotypes related to AD like amyloid burden in the brain.

Furthermore, in complex diseases like AD, the disease risk is conveyed by a number of SNPs each one with small effect on disease susceptibility, which led to the formulation of polygenic risk scores (PRS) as a successful approach in AD risk prediction (de Rojas et al., 2021). However, the limitations of PRS are the arbitrary significance thresholds for SNP selection as well as multiple testing issues (Baker and Escott-Price, 2020). In addition, PRS accounts only for the independent risk effects across multiple loci and ignores epistasis, the potentially synergistic interaction effect between genetic variants (Cordell, 2002). However, existing strategies to explore epistasis in diseases like AD are not powerful enough due to the large number of statistical tests that need to be performed (Raghavan and Tosto, 2017).

Based on these considerations, it may be hypothesized that potentially relatively small effects of single AQP4 SNPs may add up to a relevant impact in their combination. Such complex synergistic/interactive effects may be better detected by means of machine learning (ML) models as compared to conventional approaches. In contrast to the linear models used in GWAS, ML models based on decision trees like random forests or gradient boosted trees have an inherent ability to detect interactions between features, making them an excellent candidate for detecting SNP-SNP interactions even beyond only pairwise interactions (Johnsen et al., 2021).

Recently, a method of explainable artificial intelligence (AI) named "SHapley Additive exPlanation" (SHAP) has been introduced which computes explanations of tree-based ML models. It is based on game theory and has the ability to directly measure local feature interaction effects (Lundberg et al., 2020). SHAP provides also tools for understanding the global model structure based on combining the local explanations, making this method ideal for evaluating multiple SNP combinations and their importance and effects.

In this study, we tested the hypothesis that *AQP4* SNPs show a synergistic effect on brain amyloid burden. We used a parallel tree boosting method, XGBoost ("eXtreme Gradient Boosting") (Chen and Guestrin, 2016) with the minor allele carrier status of *AQP4* SNPs as features, to model interactions between SNPs as well as their individual effects on brain amyloid burden, with data from 324 participants of the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni. usc.edu). We used explanations derived from SHAP to formulate a SNP-based risk score, considering protective as well as deleterious effects of the SNPs, and evaluated the risk score using a causal modelling approach (Arnold et al., 2020) based on multiple linear regression both in ADNI as well as in an independent cohort, data from 2987 cognitively unimpaired participants of the screening process of the A4 study (a4study.org) (Sperling et al., 2014).

2. Materials and methods

Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The primary goal of ADNI was to test whether neuroimaging (serial magnetic resonance imaging (MRI), positron emission tomography (PET)) together with biological markers, clinical and neuropsychological assessment can be combined to measure the progression from cognitively normal (CN) to mild cognitive impairment (MCI) and Alzheimer's disease (AD). In 2013, whole genome sequencing (WGS) was performed on 818 participants (Alzheimer's Disease Neuroimaging Initiative (ADNI), 2013).

Further, we used data of the Anti-Amyloid Treatment in Asymptomatic Alzheimer's (A4) study (a4study.org). The A4 study is a secondary prevention trial in preclinical Alzheimer's disease, aiming to slow cognitive decline associated with brain amyloid accumulation in clinically normal older individuals (Sperling et al., 2014). Cognitively unimpaired older individuals were selected as participants, based on evidence of amyloid accumulation on screening PET scans. Participants without evidence of elevated amyloid accumulation were eligible to screen for the Longitudinal Evaluation of Amyloid Risk and Neurodegeneration (LEARN) Study. This companion observational study serves as a longitudinal comparison group with the treatment and placebo Aβ+ arms randomized in the A4 study (Sperling et al., 2020). We used the A4/LEARN pre-randomization data acquired in the screening process with imputed genetic data as from November 23, 2020.

2.1. Study participants

From the ADNI cohort, we considered only participants with whole genome sequencing (WGS) data of *AQP4* SNPs as well as completed [^{18}F]Florbetapir PET scans available. Furthermore, to ensure an ethnically homogeneous group, we excluded 10 (5 female, 5 male, no significant difference in age) out of 334 participants with an ethnic category different to "Not Hispanic or Latino".

Depending on the period since their enrollment, ADNI participants had several follow-up visits. At the time of enrollment, the diagnosis was not exceeding mild or early forms of dementia. Therefore, we decided to use data from the last visit available, to ensure that we consider the maximum bandwidth of disease from cognitively normal to severe forms of AD. Consequently, participants were older on average and were more

often diagnosed with AD compared to baseline. Overall, N= 324 participants were included.

From the A4/LEARN cohort, we also considered only participants with data of *AQP4* SNPs available, and a completed [¹⁸F]Florbetapir PET scan. We also restricted the ethnicity to "Not Hispanic or Latino" by excluding 20 (10 female, 10 male, no significant difference in age) participants with an ethnic category different to "Not Hispanic or Latino". Overall, N= 2987 participants were included.

For the assessment of significant differences in age of the two cohorts we performed an independent samples t-test using SciPy (https://scipy.org). For the assessment of significant differences in the percentage of APOE ε4 carriers and Aβ positives we used a chi-square test of independence of variables, also using SciPy.

2.2. AQP4 SNPs

Candidate *AQP4* SNPs were collected from the database of genetic variation of the National Center for Biotechnology Information (NCBI), dbSNP (Sherry et al., 2001) (https://www.ncbi.nlm.nih.gov/snp/). 31 *AQP4* SNPs were selected based on their availability in both ADNI and A4/LEARN after quality control procedures which included a minor allele frequency of greater than 5 %, and Hardy-Weinberg equilibrium ($p < 0.001$), using PLINK 2.0 (Chang et al., 2015) (https://www.cog-genomics.org/plink/2.0). We identified 3 groups of SNPs with high correlation (Pearson correlation coefficient, $r > 0.96$) and kept one SNP as representative for a group which either had a previously reported association in the context of AD or had the highest correlations with all other group members. Finally, 19 SNPs were used for the analysis. The decision to keep correlated SNPs up to an r^2 of 0.92 was based on evidence that SNP-SNP pairs in high linkage disequilibrium can be significantly associated with phenotypes due to epistatic effects (Singhal et al., 2023), so that cases where two correlated SNPs differ from each other might contain valuable information. Details of the SNPs which were used, as well as previously reported associations, are shown in Table 1. Fig. 1 shows an illustration of the position of the SNPs within the AQP4 gene. With our SNP selection we cover the entire locus of the AQP4 gene.

Finally, we assigned a "carrier" status to participants which are either homozygous in the minor allele or heterozygous, and a "non-carrier" status to participants which are homozygous in the common allele.

2.3. Amyloid PET imaging

For the ADNI cohort, brain amyloid burden was assessed using a region of interest based analysis of [¹⁸F]Florbetapir (formerly AV45) PET imaging data, which is available in a spreadsheet format from ADNI (Jagust et al., 2015). In short, after the injection of approximately 10 mCi as an intravenous bolus, participants are scanned from 50 to 70 min after injection in time frames of 5 min each. Quality control and preprocessing of the imaging data included visual inspection, correction for motion between the time frames, averaging of the motion corrected time frames to a single frame, rigid coregistration to a standard grid and intensity normalization using a reference region. To account for the various PET scanners used in the study, data were homogenized using a scanner-specific smoothing kernel. From the available spreadsheet, we used the summary [¹⁸F]Florbetapir cortical standardized uptake value ratios (SUVR) normalized by whole cerebellum. From these values, global Aβ positivity in individual participants was determined by applying a cutoff value of 1.11, as described in the internal ADNI "Amyloid PET Processing Methods" document.

For the assessment of brain amyloid burden in the A4/LEARN study, participants also received an injection of 10 mCi [¹⁸F]Florbetapir and were scanned 50–70 min after injection. Amyloid burden is also measured as mean cortical SUVR with a whole cerebellar reference region, and is available as a spreadsheed from the A4/LEARN study. To

Table 1
AQP-4 SNPs investigated in the current study.

| SNP	Consequence	Previously reported associations
rs162005	Intron	
rs9961118	2KB upstream	
rs72878794	2KB upstream	– Amyloid burden and clinical outcome in AD (Chandra et al., 2021) – Sudden infant death syndrome (Opdal et al., 2017)
rs151244	2KB upstream	– Amyloid burden and clinical outcome in AD (Chandra et al., 2021)
rs162006	2KB upstream	
rs2075575	2KB upstream	– Parkinson's disease (Sun et al., 2023) – Leukoaraiosis (Yadav et al., 2014) – Sudden infant death syndrome (Eidahl et al., 2021; Opdal et al., 2010)
rs72878787	2KB upstream	
rs162008	5'-UTR	– Brain plasticity and learning (Woo et al., 2018) – Slow wave energy regulation in human NREM sleep (Ulv Larsen et al., 2020)
rs3875089	Intron	– Moderation of the relationship between sleep and brain Aβ-amyloid burden (Rainey-Smith et al., 2018) – Traumatic brain injury outcome (Dardiotis et al., 2014) – Cognitive decline after AD diagnosis (Burfeind et al., 2017)
rs4800773	Intron	
rs3763040	Intron	– Cognitive decline after AD diagnosis (Burfeind et al., 2017) – Cognitive performance in Parkinson's disease (Fang et al., 2022)
rs162009	Intron	– Cognitive performance (Fang et al., 2022) and brain activity (Jiang et al., 2023) in Parkinson's disease
rs35248760	Synonymous	– SNPs with high correlation: rs72878776, rs67207056, rs60565102, rs55875625, rs12968026, rs11661256(all Intron), rs1058427 (3'-UTR) – rs72878776 moderates the relationship between sleep and brain Aβ-amyloid burden (Rainey-Smith et al., 2018) – rs1058427 associated with intracerebral haemorrhage (ICH) and perihematomal oedema (PHE) volume (Appelboom et al., 2015)
rs71353406	Intron	– SNPs with high correlation: rs14393, rs3763043 (all 3'-UTR) – Moderation of the relationship between sleep and brain Aβ-amyloid burden (Rainey-Smith et al., 2018) – Traumatic brain injury outcome (Dardiotis et al., 2014) (rs3763043) – Cognitive performance in Parkinson's disease (Fang et al., 2022) (rs3763043) – Cognitive decline after AD diagnosis (Burfeind et al., 2017) – Schizophrenia in the Southern Chinese Han population (Wu et al., 2020) (rs3763043)
rs71353405	Intron	
rs335930	Intron	– Slow wave energy regulation in human NREM sleep (Ulv Larsen et al., 2020)
rs1058424	3'-UTR	– Schizophrenia in the Southern Chinese Han population (Wu et al., 2020)
rs7240333	3'-UTR	– Cognitive performance in Parkinson's disease (Fang et al., 2022)
rs16942851	500B Downstream	– SNPs with high correlation: rs455671, rs335931 (all Intron), rs335929 (3'-UTR) – Slow wave energy regulation in human NREM sleep (Ulv Larsen et al., 2020) – More rapid decline in Clinical Dementia Rating, but slower decline in Logical Memory and Digit Symbol tests performance (Burfeind et al., 2017) (rs335929)
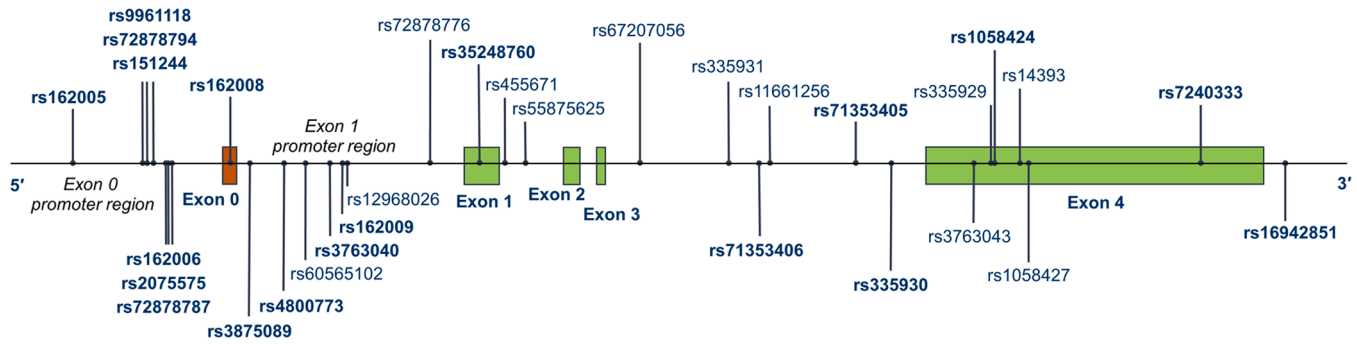
Fig. 1. Illustration of the SNP positions within the AQP4 gene. The 19 SNPs which were used for the analysis are represented in bold. The SNPs represented in regular type are in high correlation with one of the SNPs in bold (see Table 1).

maintain consistency with ADNI data, Aβ positivity was determined by applying a general cutoff value of 1.11, as opposed to the A4/LEARN procedure where SUVR values between 1.10 and 1.15 were considered as amyloid positive only when a visual read was also deemed positive by a consensus determination involving two readers (Sperling et al., 2020).

2.4. Feature engineering and estimation of risk score

Decision tree ensembles are ideal to capture complex data dependencies. To detect synergistic or interaction effects of the AQP4 SNPs we have used sex, APOE4 status expressed as number of ε4 alleles, and the status of the 19 AQP4 SNPs from ADNI as features, as well as brain amyloid burden as target for a regression task using XGBoost 1.7.6 ("**E**xtreme **G**radient **B**oosting") (Chen and Guestrin, 2016) (https://github.com/dmlc/xgboost). This algorithm uses many decision trees as weak learners and combines them to a strong learner by minimizing a loss function. It is known as a robust algorithm with excellent performance in a variety of applications. Since we did not aim for predictive modelling and model generalizability, we performed no further model tuning or hyperparameter optimization. For model evaluation we used 5-fold cross validation (scikit-learn 1.2.2) with root mean squared error (RMSE) and mean absolute error (MAE) as scoring parameters.

We have used the data from the ADNI cohort for this step because ADNI has an equal share of amyloid positive and negative participants and covers all stages of AD, from cognitively normal to dementia, in contrast to A4/LEARN with only cognitively unimpaired participants. Moreover, the A4/LEARN cohort is significantly younger than the ADNI cohort. Some participants might not yet have reached detectable amyloid levels, which would be false information for the model. Restricting the A4/LEARN cohort to participants older than 74 to pass the age of detectable amyloid onset (Bilgel et al., 2016) would result again in a small dataset, since only 27 % of the participants are older than 74.

The prevalence of Alzheimer's disease is higher in women compared to men (Niu et al., 2017). Moreover, AQP4 gene regulation has been found to be affected by sex-biasing factors (Blencowe et al., 2022), and it was also shown that AQP4 regulates the effects of ovarian hormones on neurotransmission (Sun et al., 2007). Therefore, we decided to perform the modelling separately for females and males to interrogate potential biological sex differences.

Next, we used the TreeExplainer of the SHAP package (Lundberg et al., 2020; Lundberg and Lee, 2017) (https://github.com/slundberg/shap, version 0.42.0) to explain the model decisions. The TreeExplainer can explain each prediction with respect to local feature importance and feature effect towards the predicted value, starting from the average of the whole cohort, as well as feature interaction effects. Local explanations can also be combined to understand the global model structure. Since the decision tree model that we used already considers correlations and interactions of features, the explanations of the TreeExplainer refer to the net effects of the features, given all correlations and interactions.

After visual inspection of the explanations, we assigned the carrier status of an AQP4 SNP as "negative" when the global feature effect consistently had a negative SHAP value, and as "positive" when the global feature effect consistently had a positive SHAP value.

Since the model uses the average amyloid burden of the whole cohort as a starting point, from which the prediction is shifted to higher or lower values according to the feature effects, a shift towards lower values is a shift to the normal, healthy case, and of interest for increased amyloid burden is only the shift towards higher values. Therefore we finally counted the number of positive SNPs for each individual to calculate a sex-specific SNP-based risk score both in ADNI and in A4/LEARN as part of a feature engineering process where new features which are more relevant to the target can be created from existing features using domain knowledge and data analysis. This new feature represents the joint effect of the SNPs, given all correlations and statistical interactions.

2.5. Causal modelling and statistical analysis

Usually, machine learning modelling is performed as prediction modelling which seeks to accurately predict the outcome of interest from a set of given features, preferably with high predictive performance. However, good features for predictive modelling are often rather symptoms of the outcome than causes. For example, James et al. (James et al., 2021) have found that ML models outperform existing dementia risk prediction models with only six key features, namely level of independence, clinical judgement of various aspects of decline, time to complete trail making test part B, clinical dementia rating (CDR) home and hobbies impairment, CDR memory and CDR orientation. These features have a strong predictive potential for dementia but cannot be considered as causal for the disease.

In contrast to predictive modelling, the goal of causal modelling is to estimate the true causal association between a particular variable and the outcome of interest. While randomized control trials are the gold standard to assess causal associations, they also can be studied to a certain extent in observational data by the formulation of a causal model of confounding and a careful consideration of the covariates based on this model (Arnold et al., 2020; Pearl and Mackenzie, 2018).

To represent our assumptions regarding the relationship between AQP4 status expressed as SNP-based risk score, brain amyloid burden measured as [^{18}F]Florbetapir SUVR, and known demographic and genetic risk factors (age, sex, and APOE4 status), we developed a graphical causal model as a directed acyclic graph (DAG) (Pearl, 1995) using DAGitty (http://dagitty.net/) (Fig. 2). This causal model allowed us to assess which of the covariates have to be controlled for in a regression model. AQP4 status is the exposure of interest, brain amyloid burden the outcome. APOE4 status, age and sex affect brain amyloid burden directly, which includes all unknown or unobserved mediators. We also cannot exclude a possible effect of age and sex on both AQP4 and APOE4 expression. In Chandra et al. (2021), no relationship between AQP4
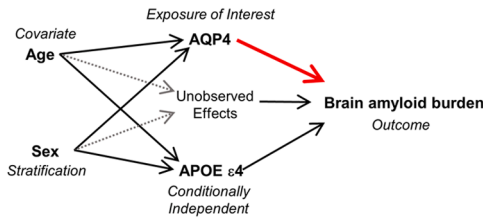
Fig. 2. Causal model as directed acyclic graph (DAG), to assess exposure-outcome adjustment set.

SNPs and APOE4 status could be established. Our own correlation analysis did also show no relationship, so that we did not include a relation between AQP4 and APOE ε4 in Fig. 2.

According to this causal model, the minimal sufficient adjustment set for estimating the total effect of AQP4 status, expressed as SNP-based risk score, on brain amyloid is age and sex. AQP4 status is then conditionally independent from APOE4 status, so that we don't have to control for it.

We stratified the sample on sex by analyzing the data separately within groups of males and females. Finally, we included age as a co-variate in a multiple linear regression model using ordinary least squares (OLS) regression within statsmodels (https://www.statsmodels.org) for both ADNI and A4/LEARN data.

To assess an association of the risk score with the prevalence of amyloid positivity, we grouped the study participants according to their risk score and compared the groups with respect to the percentage of amyloid-positive cases in the groups using the Cochran–Armitage trend test of statsmodels. To compare amyloid burden between groups of the same risk score, we performed an independent samples t-test using SciPy (https://scipy.org).

For all statistical tests we used $p < 0.05$ as the threshold for statistical significance.

3. Results

3.1. Study demographics

Table 2 shows demographic information and further characteristics for the two cohorts. The A4/LEARN cohort is significantly younger on average, and has no participants with MCI or AD. It has also significantly less APOE ε4 carriers and Aβ positives.

3.2. Feature engineering and estimation of risk score

For females, the average RMSE after performing 5-fold cross validation is 0.32 without APOE4 and 0.25 with APOE4 as feature, as well as 0.30 without APOE4 and 0.27 with APOE as feature for males. The average MAE is 0.26 (without APOE4) and 0.21 (with APOE4) for

Table 2
Demographic information and clinical characteristics of study participants.

	ADNI	A4/LEARN	p value
Number of participants	324	2987	
Age, Mean (SD) [y]	76.06 (6.98)	71.36 (4.75)	**7E−56**
Males / Females	183 / 141	1206 / 1781	
Diagnosis: Cognitively unimpaired	125	2987	
Diagnosis: Mild cognitive impairment	132	-	
Diagnosis: Dementia	64	-	
Diagnosis: non specified	3	-	
% APOE ε4 carriers	39.5	35.8	**0.0008**
% Aβ positive	50.6	32.6	**1E−10**

females as well as 0.24 (without APOE4) and 0.21 (with APOE4) for males. We therefore decided to include APOE4 into the feature set.

Fig. 3 shows the explanations for the global model structure as estimated by the TreeExplainer of SHAP for both females (left) and males (right) for ADNI data. When the model predicts the amyloid burden, it starts from the average amyloid burden of the cohort and uses the features to shift the value towards higher or lower values. The SHAP value on the x-axis represents this impact on the prediction of the model. The ordering of the features from top to bottom reflects the feature importance. As could be expected, APOE4 status is explained as the most important feature. Non-carrier status of the ε4 allele (blue dots) leads to a shift to lower predicted amyloid burden, and homozygous status of ε4 (red dots) leads to a shift to higher predicted amyloid burden than heterozygous status of ε4 (purple dots).

The explainer also shows for several SNPs that the feature is used by the model consistently to shift the prediction to either lower or higher predicted values, espressed as negative or positive SHAP values. This is especially clear for 9 SNPs in females and 7 SNPs in males. A further criterion to consider SNPs was the importance of the feature to the model, expressed by the ordering of the SNPs in the explainer. Consequently, we assigned for the blue underlined SNPs in Fig. 3 (rs335930, rs151244, rs2075575, rs35248760 and rs3875089 for women, rs3763040, rs151244, rs16942851 and rs3875089 for men) the non-carrier status as positive, in the sense that amyloid burden is increased, for the red underlined SNPs (rs71353406, rs162009, rs72878787 and rs16942851 for women, rs2075575, rs335930 and rs162009 for men) the carrier status.

There were no cases in which the selection of SNPs in an individual were all either negative or positive, so that counting the number of SNPs with a positive effect on predicted amyloid burden for each individual resulted in risk scores between 2 and 7 for females in the ADNI cohort and between 1 and 7 in the A4/LEARN cohort, and between 2 and 6 for males in ADNI as well as A4/LEARN. Fig. 4 shows that the new feature "AQP4 SNP-based risk score" is relevant to the target for both the ADNI and the A4/LEARN cohort. For both cohorts and both males and females, the t-test shows that the differences in amyloid burden between the groups of lowest and highest risk score are significant (ADNI females: p = 0.0018, males: p = 0.035; A4/LEARN females: p = 0.043, males: p = 0.035).

3.3. Causal modelling and statistical analysis

3.3.1. Association between AQP4 SNP-based risk score and brain amyloid burden

For females, multiple linear regression with age as covariate showed a significant association of risk score with amyloid burden for the ADNI (p = 0.001, regression coefficient = 0.0473, 95 % confidence interval: 0.02–0.074) as well as for the A4/LEARN (p = 0.014, regression coefficient = 0.011, 95 % confidence interval: 0.002–0.02) cohort. For males, a significant association could be shown in ADNI data (p = 0.005, regression coefficient = 0.0552, 95 % confidence interval: 0.017–0.094) but not in A4/LEARN data (p = 0.102, regression coefficient = 0.0106, 95 % confidence interval: −0.002–0.023). However, for both females and males in ADNI as well as in A4/LEARN, the mean amyloid burden differs significantly between the groups with the lowest and highest risk factors (ADNI: females p=0.002, males p=0.035; A4/LEARN: females p=0.043, males p=0.035).

3.3.2. Association between AQP4 SNP-based risk score and amyloid positivity

The distribution of PET SUVRs for the different risk scores in Fig. 4 suggests that the increase in average amyloid burden is not driven by a general, small increase for all participants with this risk score, but rather by a higher number of participants with considerably elevated amyloid burden, compared to the groups with lower risk score. This should reflect in an increase in the share of amyloid positives with increasing
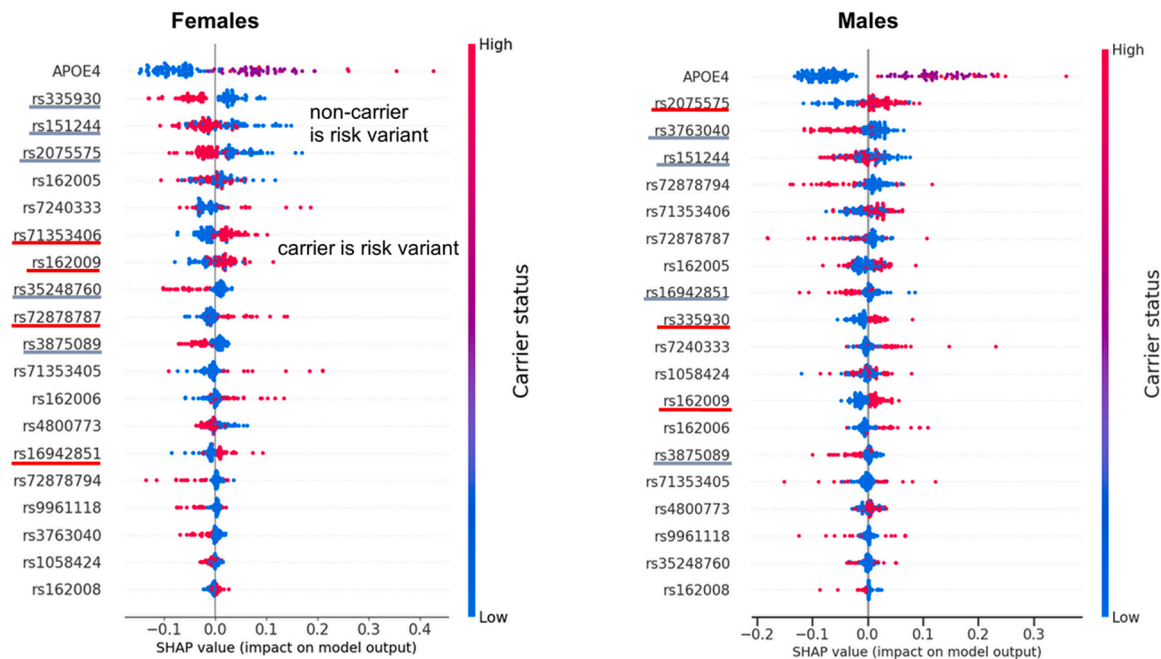
Fig. 3. The SHAP summary plot explains the global structure of the decision tree ensemble model for females (left) and males (right). SHAP values on the x-axis represent the impact of a feature on the model's prediction. Each dot of the beeswarm plot corresponds to an individual in the ADNI dataset. For AQP4 SNPs, red dots represent carriers (carrier status "high") and blue dots represent non-carriers (carrier status "low") of the SNPs. For the APOE4 feature, the colour of the dots represent the number of the ε4 alleles (0/1/2). The SNP-based risk score is calculated from the SNPs which are underlined in blue and red, so that for the blue underlined SNPs the non-carrier is the risk variant and for the red underlined SNPs the carrier is considered as the risk variant.

risk score.

Fig. 5 shows that the risk score is indeed predictive for the prevalence of amyloid positivity, with a trend towards higher prevalence with increasing risk score. The Cochran-Armitage trend tests show that the trends are significant for both cohorts in females (ADNI: $p = 0.002$, A4/LEARN: $p = 0.04$) as well as in males (ADNI: 0.0007, A4/LEARN: 0.03).

4. Discussion

In this work, we have used decision tree ensembles and explainable AI to assess the impact of interacting *AQP4* polymorphisms on brain amyloid burden. This allowed us to define a sex-specific group of SNPs with increasing effects with respect to brain amyloid burden, which together showed a highly significant association with brain amyloid burden. In particular, the increase in mean amyloid burden with higher risk scores is driven by a higher number of individuals with increased amyloid burden, which reflects in an association with amyloid positivity. Of note, the risk score which was estimated based on ADNI data was also significant in the completely independent A4/LEARN cohort of cognitively unimpaired participants with respect to both amyloid burden and share of amyloid positivity for females, and with respect to the share of amyloid positivity also for males.

For some of the individual SNPs which make up the risk score, effects on brain amyloid burden, cognition and functional decline in AD have already been described in previous work. Chandra et al. (2021) found that the carrier status of rs151244 was associated with increased Aβ uptake, but only for Aβ positive patients with late MCI or AD. Burfeind et al. (2017) found that the carrier status of rs3875089 was associated with slower cognitive decline or dementia progression. They also found that rs3763040 and rs3763043, which is in strong linkage disequilibrium (LD) with rs71353406, were associated with faster cognitive decline or dementia progression.

However, they found also an inconsistent association of rs335929, which is in strong LD with rs16942851, with slower cognitive but faster functional decline. Similarly, Fang et al. (2022) found further

inconsistent associations of rs3763040 and rs3763043 with worse executive but better visuospatial function in a cohort of PD patients. The reason for these discrepant findings might be the occurance of a statistical peculiarity named Simpson's paradox, which describes the phenomenon that a trend observed in the overall data can disappear or even reverse when the data are stratified or analyzed with respect to subgroups. This effect is the consequence of an extremely unequal distribution of the combination of the carrier or non-carrier states for two SNPs (Fig. 6). Burfeind et al. (2017) describe such a situation for rs3763040 and rs3763043. They observed that minor allele carriers of rs3763040 were most often also minor allele carriers of rs3763043, without the two SNPs being in linkage disequilibrium. A trend for one SNP might therefore be driven by the trend of the second SNP and the distribution of the carrier/non-carrier combinations of both. The discrepant findings for the effect direction (protective and deleterious) might therefore be a consequence of the presence of a second SNP with a rare carrier/non-carrier combination which is not adequately covered by the available data. Consequently, considering the joint effects of the AQP4 SNPs has the potential to yield more reliable associations.

Besides effects for amyloid burden and disease progression in AD, SNPs of the risk score have also been found to be associated with sleep measures. rs335930 and rs16942851 are part of a haplotype which is associated with slow wave energy regulation in human NREM sleep (Ulv Larsen et al., 2020). The SNP rs35248760 is in high LD with rs72878776, which has been found to be associated with poorer overall sleep quality, as well as to moderate the effect of sleep duration on brain amyloid burden (Rainey-Smith et al., 2018). This provides a direct link between brain amyloid aggregation, sleep and AQP4. However, further studies are needed to understand the underlying molecular mechanisms.

While the studies of Burfeind et al. (2017) and Chandra et al. (2021) have found associations of genetic variations of *AQP4* with measures of dementia using classical linear models, these associations were mainly noted in cohorts of patients with significant clinical symptoms. With our method, which uses the joint effect of SNPs together with a stratification for sex, we were able to find associations which were consistent and
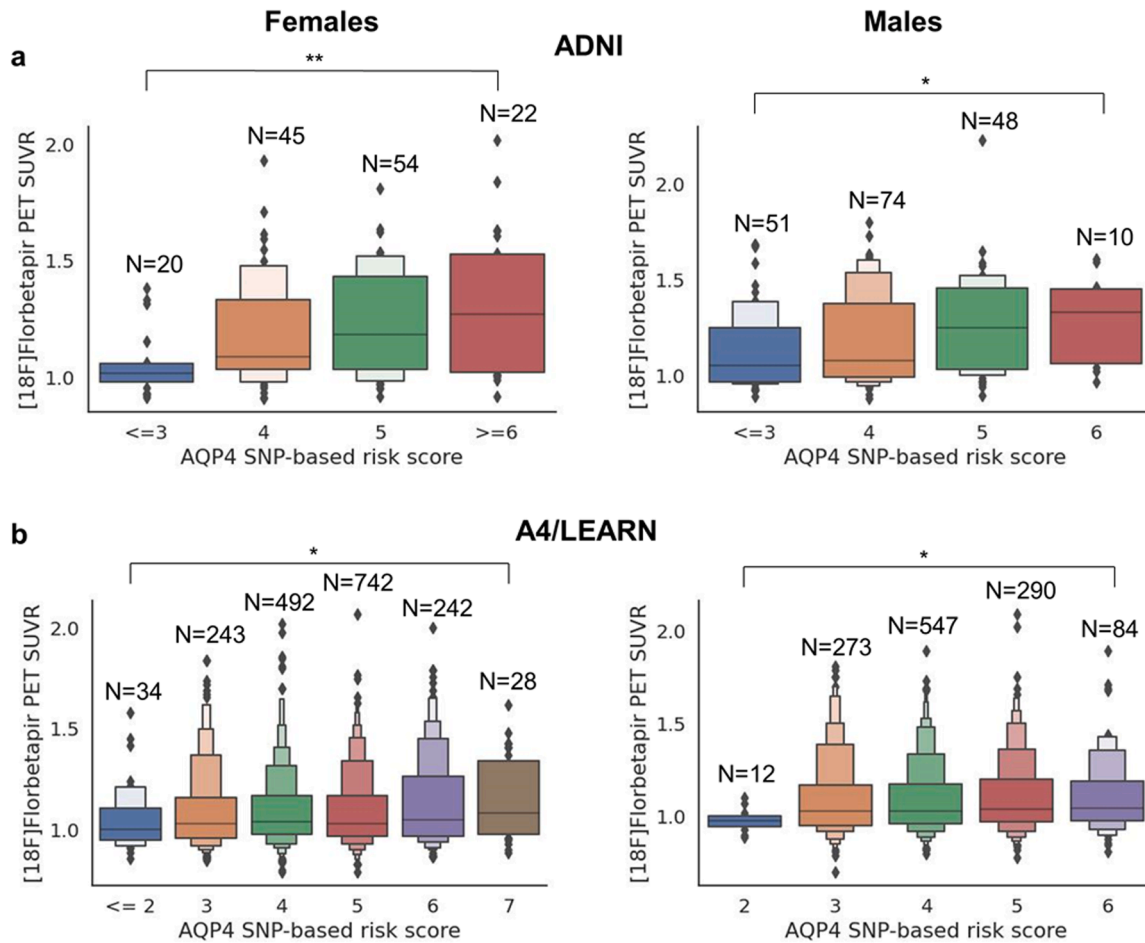
Fig. 4. (a) In the ADNI cohort the new feature "AQP4 SNP-based risk score", calculated as the sum of deleterious variants, is clearly relevant to the target, amyloid burden measured as [^{18} F] Florbetapir PET SUVR, for both females and males. (b) In the A4/LEARN cohort, an increase in amyloid burden can also be observed with increasing risk score. The increase is especially driven by the extreme low and high risk scores, and is more pronounced for females.

reproducible in a cognitively unimpaired cohort. This may provide valuable insights into the disease's underlying mechanisms at a very early stage in disease development, which opens up possibilities for early intervention or, potentially, prevention.

The joint effect of SNPs may be elucidated by considering their possible role as splicing enhancers or silencers in the transcription of AQP4 isoforms, providing a biological rationale for our method. Isoforms are protein variants that differ in the number of exons included during the transcription of genetic information. AQP4 can be translated by alternative splicing into the isoform AQP4-M1, consisting of exons 0–4 (see Fig. 1), which is freely mobile in the plasma membrane, or into the smaller isoform AQP4-M23, consisting of exons 1-4, which forms arrays in the astrocytic endfeet (Fang et al., 2022; Smith et al., 2014). Recently a further isoform was described, an elongated variant named AQP4X, which is generated by a stop codon readthrough event and which is exclusively perivascular (Sapkota et al., 2022).

A limitation of our study is the small size of the dataset which was used for modelling and explanations, since rare combinations of SNP states could be missed. We used the smaller ADNI cohort instead of the much larger A4/LEARN cohort for modelling and explanations because, especially for the later visits, all diagnoses were present, from cognitively normal to AD, while in A4/LEARN only cognitively normal participants were included. The small size of the dataset could also be a reason that we see for three of the SNPs (rs2075575, rs16942851 and rs335930) opposed effects between the sexes. From Fig. 3, it can be noted that in females the number of carriers and non-carriers of these SNPs are more balanced than in males, which could lead to the opposed

effects. A replication of our method with a larger dataset, also containing the whole bandwidth of disease, would therefore be of high interest.

A further limitation is that the results of our study might not be translated to the general population since we restricted the ethnicity to "Not Hispanic or Latino" to ensure an ethnically homogeneous group. Genetic variants might occur at different frequencies in different populations, so that rare combinations of variants in one population might be more frequent in another population. A replication of our study using datasets that contain adequate numbers of participants with other ethnicities would be of high interest.

Our method also does not differentiate between epistatic and statistical interactions. Epistatic interactions describe a masking effect, whereby a status of a variant can be dominant over another variant in the sense that it prevents the other variant from manifesting its effect (Cordell, 2002). Statistical interactions originate from the unequal distribution of the combinations of the carrier or non-carrier states of variants. The incorporation of a SNP-SNP-interaction specific model with consideration of epistatic interactions into the formulation of the risk score could have a potential to improve the performance of the risk score, and should be a topic of future investigations.

In summary, our results support the hypothesis of an involvement of the glymphatic system, and particularly AQP4, in the pathologic accumulation of ß-amyloid in the brain. Our findings suggest also that several *AQP4* SNPs show a synergistic and sex-dependent effect on early brain amyloid aggregation pathology, which can be expressed by an AQP4 polymorphism-based risk score. The risk score has the potential for practical application in stratification, such as in intervention studies or
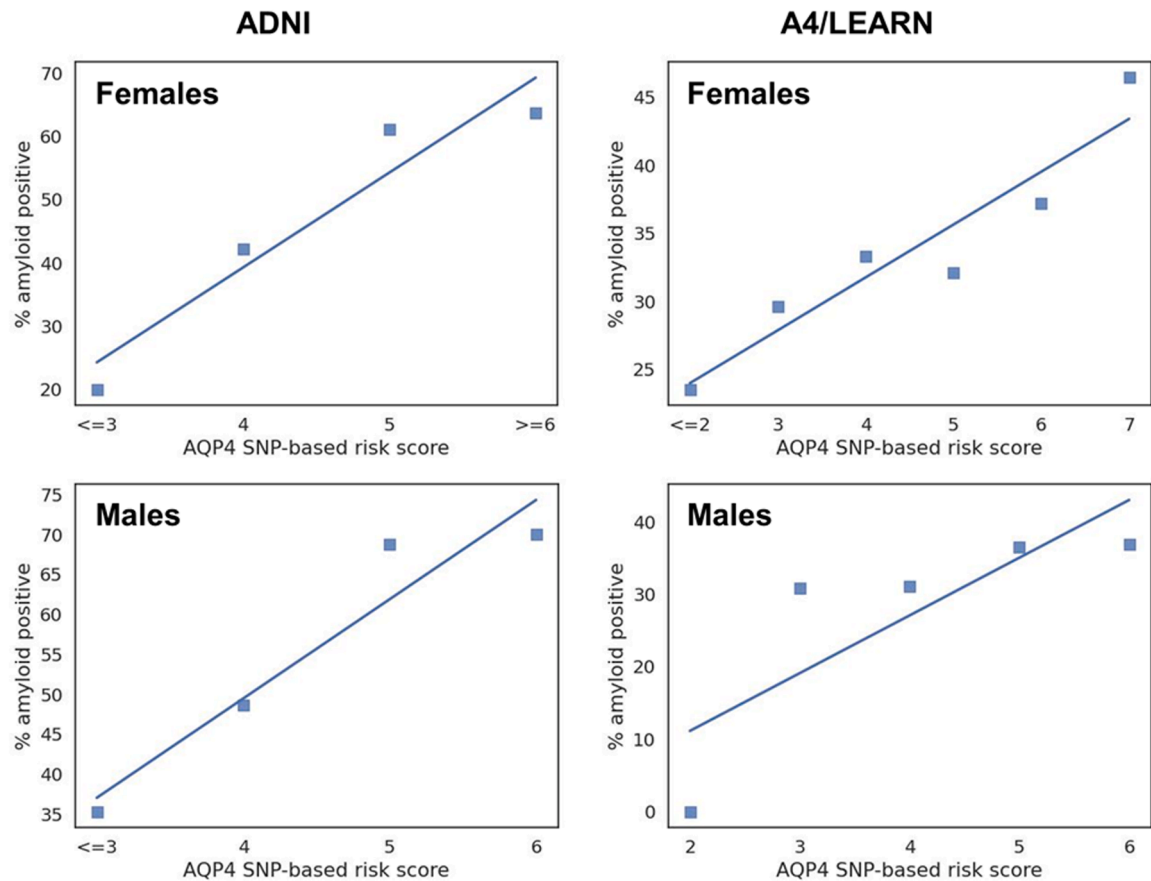
Fig. 5. The share of amyloid positives increases with higher risk scores for both males and females in the ADNI as well as in the A4/LEARN cohort.
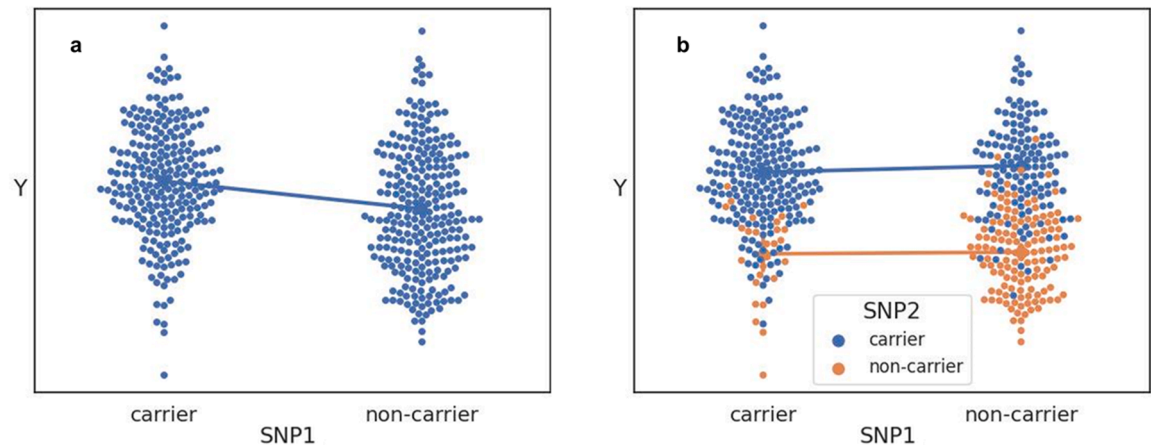

Fig. 6. Simpson's paradox: The trend for SNP1 in (a) disappears and even slightly inverses when the data are analysed in consideration of the status of SNP2 (b). This effect is a consequence of the extreme low amount of SNP2 non-carriers (orange dots) in SNP1 carriers. The two SNPs are not in linkage disequilibrium since SNP1 non-carriers can be both SNP2 carriers and non-carriers.

clinical trial recruitment. Additionally, it will be valuable for further investigations into associations with neuroinflammation and sleep, as well as into the role of AQP4 in amyloid aggregation.

CRediT authorship contribution statement

Andreas Bauer: Writing – review & editing, Conceptualization. **Alexander Drzezga:** Writing – review & editing, Supervision, Conceptualization. **Simone Beer:** Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation,

Conceptualization. **David Elmenhorst:** Writing – review & editing, Conceptualization. **Gerard Bischof:** Writing – review & editing, Conceptualization. **Alfredo Ramirez:** Writing – review & editing, Conceptualization.

Declaration of Competing Interest

None.

Acknowledgments

References

Alzheimer's Disease Neuroimaging Initiative (ADNI). 2013. Introduction and Procedures for Accessing Data from Whole Genome Sequencing of ADNI Subjects. http://adni.loni.usc.edu/wp-content/uploads/2010/09/ADNI_WGS_Notice_20130917.pdf.

Appelboom, G., Bruce, S., Duren, A., Piazza, M., Monahan, A., Christophe, B., Zoller, S., LoPresti, M., Connolly, E.S., 2015. Aquaporin-4 gene variant independently associated with oedema after intracerebral haemorrhage. Neurol. Res. 37 (8), 657–661. https://doi.org/10.1179/1743132815Y.0000000047.

Arnold, K.F., Davies, V., de Kamps, M., Tennant, P.W.G., Mbotwa, J., Gilthorpe, M.S., 2020. Reflection on modern methods: generalized linear models for prognosis and intervention—theory, practice and implications for machine learning. Int. J. Epidemiol. 49 (6), 2074–2082. https://doi.org/10.1093/ije/dyaa049.

Baker, E., Escott-Price, V., 2020. Polygenic risk scores in Alzheimer's disease: current applications and future directions. Front. Digit. Health 2, 14. https://doi.org/10.3389/fdgth.2020.00014.

Bellenguez, C., Küçükali, F., Jansen, I.E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A.C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., Holmans, P.A., Boland, A., Damotte, V., van der Lee, S.J., Costa, M.R., Kuulasmaa, T., Yang, Q., de Rojas, I., Bis, J.C., Yaqub, A., Prokic, I., Chapuis, J., Ahmad, S., Giedraitis, V., Aarsland, D., Garcia-Gonzalez, P., Abdelnour, C., Alarcón-Martín, E., Alcolea, D., Alegret, M., Alvarez, I., Álvarez, V., Armstrong, N.J., Tsolaki, A., Antúnez, C., Appollonio, I., Arcaro, M., Archetti, S., Pastor, A.A., Arosio, B., Athanasiu, L., Bailly, H., Banaj, N., Baquero, M., Barral, S., Beiser, A., Pastor, A.B., Below, J.E., Benchek, P., Benussi, L., Berr, C., Besse, C., Bessi, V., Binetti, G., Bizarro, A., Blesa, R., Boada, M., Boerwinkle, E., Borroni, B., Boschi, S., Bossù, P., Bråthen, G., Bressler, J., Bresner, C., Brodaty, H., Brookes, K.J., Brusco, L.I., Buiza-Rueda, D., Bûrger, K., Burholt, V., Bush, W.S., Calero, M., Cantwell, L.B., Chene, G., Chung, J., Cuccaro, M.L., Carracedo, Á., Cecchetti, R., Cervera-Carles, L., Charbonnier, C., Chen, H.-H., Chillotti, C., Ciccone, S., Claassen, J.A.H.R., Clark, C., Conti, E., Corma-Gómez, A., Costantini, E., Custodero, C., Daian, D., Dalmasso, M.C., Daniele, A., Dardiotis, E., Dartigues, J.-F., de Deyn, P.P., de Paiva Lopes, K., de Witte, L.D., Debette, S., Deckert, J., del Ser, T., Denning, N., DeStefano, A., Dichgans, M., Diehl-Schmid, J., Diez-Fairen, M., Rossi, P.D., Djurovic, S., Duron, E., Düzel, E., Dufouil, C., Eiriksdottir, G., Engelborghs, S., Escott-Price, V., Espinosa, A., Ewers, M., Faber, K. M., Fabrizio, T., Nielsen, S.F., Fardo, D.W., Farotti, L., Fenoglio, C., Fernández-Fuertes, M., Ferrari, R., Ferreira, C.B., Ferri, E., Fin, B., Fischer, P., Fladby, T., Fließbach, K., Fongang, B., Fornage, M., Fortea, J., Foroud, T.M., Fostinelli, S., Fox, N.C., Franco-Macías, E., Bullido, M.J., Frank-García, A., Froelich, L., Fulton-Howard, B., Galimberti, D., García-Alberca, J.M., García-González, P., Garcia-Madrona, S., Garcia-Ribas, G., Ghidoni, R., Giegling, I., Giorgio, G., Goate, A.M., Goldhardt, O., Gomez-Fonseca, D., González-Pérez, A., Graff, C., Grande, G., Green, E., Grimmer, T., Grünblatt, E., Grunin, M., Gudnason, V., Guetta-Baranes, T., Haapasalo, A., Hadjigeorgiou, G., Haines, J.L., Hamilton-Nelson, K.L., Hampel, H., Hanon, O., Hardy, J., Hartmann, A.M., Hausner, L., Harwood, J., Heilmann-Heimbach, S., Helisalmi, S., Heneka, M.T., Hernández, I., Herrmann, M.J., Hoffmann, P., Holmes, C., Holstege, H., Vilas, R.H., Hulsman, M., Humphrey, J., Biessels, G.J., Jian, X., Johansson, C., Jun, G.R., Kastumata, Y., Kauwe, J., Kehoe, P. G., Kilander, L., Ståhlbom, A.K., Kivipelto, M., Koivisto, A., Kornhuber, J., Kosmidis, M.H., Kukull, W.A., Kuksa, P.P., Kunkle, B.W., Kuzma, A.B., Lage, C., Laukka, E.J., Launer, L., Lauria, A., Lee, C.-Y., Lehtisalo, J., Lerch, O., Lleó, A., Longstreth, W., Lopez, O., de Munain, A.L., Love, S., Löwemark, M., Luckcuck, L., Lunetta, K.L., Ma, Y., Macías, J., MacLeod, C.A., Maier, W., Mangialasche, F., Spallazzi, M., Marquié, M., Marshall, R., Martin, E.R., Montes, A.M., Rodríguez, C. M., Masullo, C., Mayeux, R., Mead, S., Mecocci, P., Medina, M., Meggy, A., Mehrabian, S., Mendoza, S., Menéndez-González, M., Mir, P., Moebus, S., Mol, M., Molina-Porcel, L., Montrreal, L., Morelli, L., Moreno, F., Morgan, K., Mosley, T., Nöthen, M.M., Muchnik, C., Mukherjee, S., Nacmias, B., Ngandu, T., Nicolas, G., Nordestgaard, B.G., Olaso, R., Orellana, A., Orsini, M., Ortega, G., Padovani, A., Paolo, C., Papenberg, G., Parnetti, L., Pasquier, F., Pastor, P., Peloso, G., Pérez-Cordón, A., Pérez-Tur, J., Pericard, P., Peters, O., Pijnenburg, Y.A.L., Pineda, J.A., Piñol-Ripoll, G., Pisanu, C., Polak, T., Popp, J., Posthuma, D., Priller, J., Puerta, R., Quenez, O., Quintela, I., Thomassen, J.Q., Rábano, A., Rainero, I., Rajabli, F., Ramakers, I., Real, L.M., Reinders, M.J.T., Reitz, C., Reyes-Dumeyer, D., Ridge, P., Riedel-Heller, S., Riederer, P., Roberto, N., Rodriguez-Rodriguez, E., Rongve, A., Allende, I.R., Rosende-Roca, M., Royo, J.L., Rubino, E., Rujescu, D., Sáez, M.E., Sakka, P., Saltvedt, I., Sanabria, Á., Sánchez-Arjona, M.B., 2022. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat. Genet. 54 (4), 412–436. https://doi.org/10.1038/s41588-022-01024-z.

Bilgel, M., An, Y., Zhou, Y., Wong, D.F., Prince, J.L., Ferrucci, L., Resnick, S.M., 2016. Individual estimates of age at detectable amyloid onset for risk factor assessment. Alzheimer'S. Dement. 12 (4), 373–379. https://doi.org/10.1016/j.jalz.2015.08.166.

Blencowe, M., Chen, X., Zhao, Y., Itoh, Y., McQuillen, C.N., Han, Y., Shou, B.L., McClusky, R., Reue, K., Arnold, A.P., 2022. Relative contributions of sex hormones, sex chromosomes, and gonads to sex differences in tissue gene regulation. Genome Res. 32 (5), 807–824. https://doi.org/10.1101/gr.275965.121.

Burfeind, K.G., Murchison, C.F., Westaway, S.K., Simon, M.J., Erten-Lyons, D., Kaye, J. A., Quinn, J.F., Iliff, J.J., 2017. The effects of noncoding aquaporin-4 single-nucleotide polymorphisms on cognition and functional progression of Alzheimer's disease. Alzheimer'S. Dement. Transl. Res. Clin. Interv. 3 (3), 348–359. https://doi.org/10.1016/j.trci.2017.05.001.

Chandra, A., Farrell, C., Wilson, H., Dervenoulas, G., De Natale, E.R., Politis, M., 2021. Aquaporin-4 polymorphisms predict amyloid burden and clinical outcome in the Alzheimer's disease spectrum. Neurobiol. Aging 97, 1–9. https://doi.org/10.1016/j.neurobiolaging.2020.06.007.

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. pp. s13742-13015-10047-13748, GigaScience 4 (1). https://doi.org/10.1186/s13742-015-0047-8.

T. Chen, C. Guestrin Xgboost: A Scalable Tree Boosting System, KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. (2016), pp. 785-794, 10.1145/2939672.2939785.

Cordell, H.J., 2002. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum. Mol. Genet. 11 (20), 2463–2468. https://doi.org/10.1093/hmg/11.20.2463.

Dardiotis, E., Paterakis, K., Tsivgoulis, G., Tsintou, M., Hadjigeorgiou, G.F., Dardioti, M., Grigoriadis, S., Simeonidou, C., Komnos, A., Kapsalaki, E., Fountas, K., Hadjigeorgiou, G.M., 2014. AQP4 tag single nucleotide polymorphisms in patients with traumatic brain injury. J. Neurotrauma 31 (23), 1920–1926. https://doi.org/10.1089/neu.2014.3347.

Eidahl, J.M.L., Stray-Pedersen, A., Rognum, T.O., Opdal, S.H., 2021. Aquaporin 4 expression in the hippocampus in sudden infant death syndrome and sudden unexplained death in childhood. J. Chem. Neuroanat. 115, 101962 https://doi.org/10.1016/j.jchemneu.2021.101962.

Fang, Y., Dai, S., Jin, C., Si, X., Gu, L., Song, Z., Gao, T., Chen, Y., Yan, Y., Yin, X., 2022. Aquaporin-4 polymorphisms are associated with cognitive performance in Parkinson's Disease. Front. Aging Neurosci. 13, 740491 https://doi.org/10.3389/fnagi.2021.740491.

Hablitz, L.M., Plá, V., Giannetto, M., Vinitsky, H.S., Stæger, F.F., Metcalfe, T., Nguyen, R., Benrais, A., Nedergaard, M., 2020. Circadian control of brain glymphatic and lymphatic fluid flow. Nat. Commun. 11 (1), 4411. https://doi.org/10.1038/s41467-020-18115-2.

Hubbard, J.A., Szu, J.I., Binder, D.K., 2018. The role of aquaporin-4 in synaptic plasticity, memory and disease. Brain Res. Bull. 136, 118–129. https://doi.org/10.1016/j.brainresbull.2017.02.011.

Iliff, J.J., Wang, M., Liao, Y., Plogg, B.A., Peng, W., Gundersen, G.A., Benveniste, H., Vates, G.E., Deane, R., Goldman, S.A., Nagelhus, E.A., Nedergaard, M., 2012. A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β. Sci. Transl. Med. 4 (147), pp. 147ra111-147ra111, 10.1126/scitranslmed.3003748.

Jagust, W.J., Landau, S.M., Koeppe, R.A., Reiman, E.M., Chen, K., Mathis, C.A., Price, J.C., Foster, N.L., Wang, A.Y., 2015. The Alzheimer's disease neuroimaging initiative 2 PET core: 2015. Alzheimer'S. Dement. 11 (7), 757–771. https://doi.org/10.1016/j.jalz.2015.05.001.

James, C., Ranson, J.M., Everson, R., Llewellyn, D.J., 2021. Performance of machine learning algorithms for predicting progression to dementia in memory clinic patients. JAMA Netw. Open 4 (12), e2136553. https://doi.org/10.1001/jamanetworkopen.2021.36553.

Jiang, M., Fang, Y., Dai, S., Si, X., Wang, Z., Tang, J., Gao, T., Liu, Y., Song, Z., Pu, J., Zhang, B., 2023. The effects of AQP4 rs162009 on resting-state brain activity in Parkinson's disease. CNS Neurosci. Ther. 29 (9), 2645–2655. https://doi.org/10.1111/cns.14208.

Johnsen, P.V., Riemer-Sørensen, S., DeWan, A.T., Cahill, M.E., Langaas, M., 2021. A new method for exploring gene–gene and gene–environment interactions in GWAS with tree ensemble methods and SHAP values. BMC Bioinf. 22 (1), 230. https://doi.org/10.1186/s12859-021-04041-7.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67. https://doi.org/10.1038/s42256-019-0138-9.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions Advances in neural information processing systems. Adv. Neural Inf. Process. Syst. 30.

Mader, S., Brimberg, L., 2019. Aquaporin-4 water channel in the brain and its implication for health and disease. Cells 8 (2), 90. https://doi.org/10.3390/cells8020090.

Nedergaard, M., Goldman, S.A., 2020. Glymphatic failure as a final common pathway to dementia. Science 370 (6512), 50–56. https://doi.org/10.1126/science.abb8739.

Niu, H., Álvarez-Álvarez, I., Guillén-Grima, F., Aguinaga-Ontoso, I., 2017. Prevalencia e incidencia de la enfermedad de Alzheimer en Europa: metaanálisis. Neurol. ía 32 (8), 523–532. https://doi.org/10.1016/j.nrl.2016.02.016.

Opdal, S.H., Vege, Å., Stray-Pedersen, A., Rognum, T.O., 2010. Aquaporin-4 gene variation and sudden infant death syndrome. Pediatr. Res. 68 (1), 48–51. https://doi.org/10.1203/PDR.0b013e3181df4e7c.

Opdal, S.H., Vege, Å., Stray-Pedersen, A., Rognum, T.O., 2017. The gene encoding the inwardly rectifying potassium channel Kir4.1 may be involved in sudden infant death syndrome. Acta Paediatr. 106 (9), 1474–1480. https://doi.org/10.1111/apa.13928.

Pearl, J., 1995. Causal diagrams for empirical research. Biometrika 82 (4), 669–688. https://doi.org/10.1093/biomet/82.4.669.

Pearl, J., Mackenzie, D., 2018. The book of why: the new science of cause and effect. Basic books.

Raghavan, N., Tosto, G., 2017. Genetics of Alzheimer's Disease: the Importance of Polygenic and Epistatic Components. Curr. Neurol. Neurosci. Rep. 17 (10), 78. https://doi.org/10.1007/s11910-017-0787-1.

Rainey-Smith, S.R., Mazzucchelli, G.N., Villemagne, V.L., Brown, B.M., Porter, T., Weinborn, M., Bucks, R.S., Milicic, L., Sohrabi, H.R., Taddei, K., Ames, D., Maruff, P., Masters, C.L., Rowe, C.C., Salvado, O., Martins, R.N., Laws, S.M., 2018. A.R.G. for the genetic variation in Aquaporin-4 moderates the relationship between sleep and brain Aβ-amyloid burden. Transl. Psychiatry 8 (1), 47. https://doi.org/10.1038/s41398-018-0094-x.

de Rojas, I., Moreno-Grau, S., Tesi, N., Grenier-Boley, B., Andrade, V., Jansen, I.E., Pedersen, N.L., Stringa, N., Zettergren, A., Hernández, I., Montrreal, L., Antúnez, C., Antonell, A., Tankard, R.M., Bis, J.C., Sims, R., Bellenguez, C., Quintela, I., González-Perez, A., Calero, M., Franco-Macías, E., Macías, J., Blesa, R., Cervera-Carles, L., Menéndez-González, M., Frank-García, A., Royo, J.L., Moreno, F., Huerto Vilas, R., Baquero, M., Diez-Fairen, M., Lage, C., García-Madrona, S., García-González, P., Alarcón-Martín, E., Valero, S., Sotolongo-Grau, O., Ullgren, A., Naj, A.C., Lemstra, A. W., Benaque, A., Pérez-Cordón, A., Benussi, A., Rábano, A., Padovani, A., Squassina, A., de Mendonça, A., Arias Pastor, A., Kok, A.A.L., Meggy, A., Pastor, A.B., Espinosa, A., Corma-Gómez, A., Martín Montes, A., Sanabria, Á., DeStefano, A.L., Schneider, A., Haapasalo, A., Kinhult Ståhlbom, A., Tybjærg-Hansen, A., Hartmann, A.M., Spottke, A., Corbatón-Anchuelo, A., Rongve, A., Borroni, B., Arosio, B., Nacmias, B., Nordestgaard, B.G., Kunkle, B.W., Charbonnier, C., Abdelnour, C., Masullo, C., Martínez Rodríguez, C., Muñoz-Fernandez, C., Dufouil, C., Graff, C., Ferreira, C.B., Chillotti, C., Reynolds, C.A., Fenoglio, C., Van Broeckhoven, C., Clark, C., Pisanu, C., Satizabal, C.L., Holmes, C., Buiza-Rueda, D., Aarsland, D., Rujescu, D., Alcolea, D., Galimberti, D., Wallon, D., Seripa, D., Grünblatt, E., Dardiotis, E., Düzel, E., Scarpini, E., Conti, E., Rubino, E., Gelpi, E., Rodriguez-Rodriguez, E., Duron, E., Boerwinkle, E., Ferri, E., Tagliavini, F.,

Küçükali, F., Pasquier, F., Sanchez-Garcia, F., Mangialasche, F., Jessen, F., Nicolas, G., Selbæk, G., Ortega, G., Chêne, G., Hadjigeorgiou, G., Rossi, G., Spalletta, G., Giaccone, G., Grande, G., Binetti, G., Papenberg, G., Hampel, H., Bailly, H., Zetterberg, H., Soininen, H., Karlsson, I.K., Alvarez, I., Appollonio, I., Giegling, I., Skoog, I., Saltvedt, I., Rainero, I., Rosas Allende, I., Hort, J., Diehl-Schmid, J., Van Dongen, J., Vidal, J.-S., Lehtisalo, J., Wiltfang, J., Thomassen, J.Q., Kornhuber, J., Haines, J.L., Vogelsang, J., Pineda, J.A., Fortea, J., Popp, J., Deckert, J., Buerger, K., Morgan, K., Fließbach, K., Sleegers, K., Molina-Porcel, L., Kilander, L., Weinhold, L., Farrer, L.A., Wang, L.-S., Kleineidam, L., Farotti, L., Parnetti, L., Tremolizzo, L., Hausner, L., Benussi, L., Froelich, L., Ikram, M.A., Deniz-Naranjo, M.C., Tsolaki, M., Rosende-Roca, M., Löwenmark, M., Hulsman, M., Spallazzi, M., Pericak-Vance, M.A., Esiri, M., Bernal Sánchez-Arjona, M., Dalmasso, M.C., Martínez-Larrad, M.T., Arcaro, M., Nöthen, M.M., Fernández-Fuertes, M., Dichgans, M., Ingelsson, M., Herrmann, M.J., Scherer, M., Vyhnalek, M., Kosmidis, M.H., Yannakoulia, M., Schmid, M., Ewers, M., Heneka, M.T., Wagner, M., Scamosci, M., Kivipelto, M., Hiltunen, M., Zulaica, M., Alegret, M., Fornage, M., Roberto, N., van Schoor, N.M., Seidu, N.M., Banaj, N., Armstrong, N.J., Scarmeas, N., Scherbaum, N., Goldhardt, O., Hanon, O., Peters, O., Skrobot, O.A., Quenez, O., Lerch, O., Bossù, P., Caffarra, P., Dionigi Rossi, P., Sakka, P., Mecocci, P., Hoffmann, P., Holmans, P.A., Fischer, P., Riederer, P., Yang, Q., Marshall, R., Kalaria, R.N., Mayeux, R., Vandenberghe, R., Cecchetti, R., Ghidoni, R., Frikke-Schmidt, R., Sorbi, S., Hägg, S., Engelborghs, S., Helisalmi, S., Botne Sando, S., Kern, S., Archetti, S., Boschi, S., Fostinelli, S., Gil, S., Mendoza, S., Mead, S., Ciccone, S., Djurovic, S., Heilmann-Heimbach, S., Riedel-Heller, S., Kuulasmaa, T., del Ser, T., Lebouvier, T., Polak, T., Ngandu, T., Grimmer, T., Bessi, V., Escott-Price, V., Giedraitis, V., Deramecourt, V., Maier, W., Jian, X., Pijnenburg, Y.A.L., Smith, A.D., Saenz, A., Bizzarro, A., Lauria, A., Vacca, A., Solomon, A., Anastasiou, A., Richardson, A., Boland, A., Koivisto, A., Daniele, A., Greco, A., Marianthi, A., McGuinness, B., Fin, B., Ferrari, C., Custodero, C., Ferrarese, C., Ingino, C., Mangone, C., Reyes Toso, C., Martínez, C., Cuesta, C., Muchnik, C., Joachim, C., Ortiz, C., Besse, C., Johansson, C., Zoia, C.P., Laske, C., Anastasiou, C., Palacio, D.L., Politis, D.G., Janowitz, D., Craig, D., Mann, D.M., Neary, D., Jürgen, D., Daian, D., Belezhanska, D., Kohler, E., Castaño, E.M., Koutsouraki, E., Chipi, E., De Roeck, E., Costantini, E., 2021. E. contributors common variants in Alzheimer's disease and risk stratification by polygenic risk scores. Nat. Commun. 12 (1), 3417. https://doi.org/10.1038/s41467-021-22491-8.

Salman, M.M., Kitchen, P., Halsey, A., Wang, M.X., Törnroth-Horsefield, S., Conner, A.C., Badaut, J., Iliff, J.J., Bill, R.M., 2022. Emerging roles for dynamic aquaporin-4 subcellular relocalization in CNS water homeostasis. Brain 145 (1), 64–75. https://doi.org/10.1093/brain/awab311.

Sapkota, D., Florian, C., Doherty, B.M., White, K.M., Reardon, K.M., Ge, X., Garbow, J.R., Yuede, C.M., Cirrito, J.R., Dougherty, J.D., 2022. Aqp4 stop codon readthrough facilitates amyloid-β clearance from the brain. Brain 145 (9), 2982–2990. https://doi.org/10.1093/brain/awac199.

Sherry, S.T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., 2001. K. Sirotkin dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29 (1), 308–311. https://doi.org/10.1093/nar/29.1.308.

Simon, M., Wang, M.X., Ismail, O., Braun, M., Schindler, A.G., Reemmer, J., Wang, Z., Haveliwala, M.A., O'Boyle, R.P., Han, W.Y., Roese, N., Grafe, M., Woltjer, R., Boison, D., Iliff, J.J., 2022. Loss of perivascular aquaporin-4 localization impairs glymphatic exchange and promotes amyloid β plaque formation in mice. Alzheimer'S. Res. Ther. 14 (1), 59. https://doi.org/10.1186/s13195-022-00999-5.

Singhal, P., Veturi, Y., Dudek, S.M., Lucas, A., Frase, A., van Steen, K., Schrodi, S.J., Fasel, D., Weng, C., Pendergrass, R., Schaid, D.J., Kullo, I.J., Dikilitas, O., Sleiman, P. M.A., Hakonarson, H., Moore, J.H., Williams, S.M., Ritchie, M.D., Verma, S.S., 2023. Evidence of epistasis in regions of long-range linkage disequilibrium across five complex diseases in the UK Biobank and eMERGE datasets. Am. J. Hum. Genet. 110 (4), 575–591. https://doi.org/10.1016/j.ajhg.2023.03.007.

Smith, A.J., Jin, B.-J., Ratelade, J., Verkman, A.S., 2014. Aggregation state determines the localization and function of M1– and M23–aquaporin-4 in astrocytes. J. Cell Biol. 204 (4), 559–573. https://doi.org/10.1083/jcb.201308118.

Sperling, R.A., Donohue, M.C., Raman, R., Sun, C.-K., Yaari, R., Holdridge, K., Siemers, E., Johnson, K.A., Aisen, P.S., 2020. A.S.T. for the Association of Factors with Elevated Amyloid Burden in Clinically Normal Older Individuals. JAMA Neurol. 77 (6), 735–745. https://doi.org/10.1001/jamaneurol.2020.0387.

Sperling, R.A., Rentz, D.M., Johnson, K.A., Karlawish, J., Donohue, M., Salmon, D.P., Aisen, P., 2014. The A4 study: stopping AD before symptoms begin? Sci. Transl. Med. 6 (228), pp. 228fs213-228fs213.

Sun, X.-L., Ding, J.-H., Fan, Y., Zhang, J., Gao, L., Hu, G., 2007. Aquaporin 4 regulates the effects of ovarian hormones on monoamine neurotransmission. Biochem. Biophys. Res. Commun. 353 (2), 457–462. https://doi.org/10.1016/j.bbrc.2006.12.040.

Sun, X., Tian, Q., Yang, Z., Liu, Y., Li, C., Hou, B., Xie, A., 2023. Association of AQP4 single nucleotide polymorphisms (rs335929 and rs2075575) with Parkinson's disease: a case-control study. Neurosci. Lett. 797, 137062 https://doi.org/10.1016/j.neulet.2023.137062.

Ulv Larsen, S.M., Landolt, H.-P., Berger, W., Nedergaard, M., Knudsen, G.M., Holst, S.C., 2020. Haplotype of the astrocytic water channel AQP4 is associated with slow wave energy regulation in human NREM sleep. PLOS Biol. 18 (5), e3000623 https://doi.org/10.1371/journal.pbio.3000623.

Valenza, M., Facchinetti, R., Steardo, L., Scuderi, C., 2020. Altered waste disposal system in aging and Alzheimer's disease: focus on astrocytic aquaporin-4. Front. Pharmacol. 10, 1656. https://doi.org/10.3389/fphar.2019.01656.

Woo, J., Kim, J.E., Im, J.J., Lee, J., Jeong, H.S., Park, S., Jung, S.Y., An, H., Yoon, S., Lim, S.M., Lee, S., Ma, J., Shin, E.Y., Han, Y.E., Kim, B., Lee, E.H., Feng, L., Chun, H., Yoon, B.E., Kang, I., Dager, S.R., Lyoo, I.K., Lee, C.J., 2018. Astrocytic water channel aquaporin-4 modulates brain plasticity in both mice and humans: a potential

gliogenetic mechanism underlying language-associated learning. Mol. Psychiatry 23 (4), 1021–1030. https://doi.org/10.1038/mp.2017.113.

Wu, Y.-F., Sytwu, H.-K., Lung, F.-W., 2020. Polymorphisms in the human Aquaporin 4 Gene Are associated with schizophrenia in the Southern Chinese han population: a case–control study. Front. Psychiatry 11, 596. https://doi.org/10.3389/fpsyt.2020.00596.

Yadav, B.K., Oh, S.-Y., Kim, N.-K., Shin, B.-S., 2014. Association of rs2075575 and rs9951307 Polymorphisms of AQP-4 Gene with Leukoaraiosis. J. Stroke Cerebrovasc. Dis. 23 (5), 1199–1206. https://doi.org/10.1016/j.jstrokecerebrovasdis.2013.10.017.

Zeppenfeld, D.M., Simon, M., Haswell, J.D., D'Abreo, D., Murchison, C., Quinn, J.F., Grafe, M.R., Woltjer, R.L., Kaye, J., Iliff, J.J., 2017. Association of perivascular localization of Aquaporin-4 With cognition and alzheimer disease in aging brains. JAMA Neurol. 74 (1), 91–99. https://doi.org/10.1001/jamaneurol.2016.4370.