Journal of Alzheimer's Disease 97 (2024) 1807–1827
DOI 10.3233/JAD-231020
IOS Press

Disentangling Accelerated Cognitive Decline from the Normal Aging Process and Unraveling Its Genetic Components: A Neuroimaging-Based Deep Learning Approach

Yulin Dai[a,1], Yu-Chun Hsu[b,1], Brisa S. Fernandes[a], Kai Zhang[b], Xiaoyang Li[a,c], Nitesh Enduru[a,d], Andi Liu[a,d], Astrid M. Manuel[a], Xiaoqian Jiang[b,*] and Zhongming Zhao[a,d,e,*] for the Alzheimer's Disease Neuroimaging Initiative[2]

[a]*Center for Precision Health, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA*

[b]*Center for Secure Artificial Intelligence for Healthcare, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA*

[c]*Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA*

[d]*Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA*

[e]*Department of Biomedical Informatics, Vanderbilt University Medical enter, Nashville, TN, USA*

Accepted 11 December 2023
Pre-press 31 January 2024

Abstract.

Background: The progressive cognitive decline, an integral component of Alzheimer's disease (AD), unfolds in tandem with the natural aging process. Neuroimaging features have demonstrated the capacity to distinguish cognitive decline changes stemming from typical brain aging and AD between different chronological points.

Objective: To disentangle the normal aging effect from the AD-related accelerated cognitive decline and unravel its genetic components using a neuroimaging-based deep learning approach.

[1]These authors contributed equally to this work.

[2]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

*Correspondence to: Zhongming Zhao, PhD, Center for Precision Health, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, 7000 Fannin St., Suite 600, Houston, TX 77030, USA. Tel.: +1 713 500 3631; E-mail: Zhongming.Zhao@uth.tmc.edu and Xiaoqian Jiang, PhD, Center for Secure Artificial Intelligence for Healthcare, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, 7000 Fannin St., Suite 600, Houston, TX 77030, USA. Tel.: +1 713 500 3930; E-mail: Xiaoqian.Jiang@uth.tmc.edu.

ISSN 1387-2877/$35.00

Methods: We developed a deep-learning framework based on a dual-loss Siamese ResNet network to extract fine-grained information from the longitudinal structural magnetic resonance imaging (MRI) data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. We then conducted genome-wide association studies (GWAS) and post-GWAS analyses to reveal the genetic basis of AD-related accelerated cognitive decline.
Results: We used our model to process data from 1,313 individuals, training it on 414 cognitively normal people and predicting cognitive assessment for all participants. In our analysis of accelerated cognitive decline GWAS, we identified two genome-wide significant loci: *APOE* locus (chromosome 19 p13.32) and rs144614292 (chromosome 11 p15.1). Variant rs144614292 (G > T) has not been reported in previous AD GWA studies. It is within the intronic region of *NELL1*, which is expressed in neurons and plays a role in controlling cell growth and differentiation. The cell-type-specific enrichment analysis and functional enrichment of GWAS signals highlighted the microglia and immune-response pathways.
Conclusions: Our deep learning model effectively extracted relevant neuroimaging features and predicted individual cognitive decline. We reported a novel variant (rs144614292) within the NELL1 gene.

Keywords: Alzheimer's disease, cognitive decline, deep learning, genome-wide association study, neuroimaging

INTRODUCTION

Alzheimer's disease (AD) is a progressive and degenerative disease of the brain affecting the daily activities of the aging population. Approximately 6.2 million people in the US live with AD, and the number of individuals with AD is predicted to double by 2025. Cognitive decline and memory impairment are the prominent symptoms of AD [1]. Late-onset AD (LOAD) heritability is as high as 79% [2–5]. Despite the fact that the genetic architecture of LOAD has been identified using millions of participants [6, 7], currently, there is no effective treatment for preventing the development of AD [8, 9]. One of the reasons for this lack of proper identification and effective treatments is that we do not have a coherent and actionable system capable of accurately detecting AD and untangling its effects from the normal aging process. The widely-used Mini-Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) are strongly influenced by the individual status and non-cognitive domains, such as language, levels of literacy, and cultural and ethical norms [10]. Furthermore, fluctuations in the MMSE and ADAS-Cog tests might lack components sensitive to identifying early-stage dementia, especially mild cognitive impairment (MCI) [10–12], hindering accurate cognitive assessments and leading to misclassification due to test-specific biases.

Because AD is considered to primarily affect the brain, neuroimaging has emerged as a crucial tool for detecting structural changes in the brain associated with memory decline and for addressing the progression of AD [13–17]. Alteration in the hippocampus assessed by magnetic resonance imaging (MRI) can occur simultaneously with the first time amyloid deposition, as early as 18 years prior to dementia [9]. Yet, previous neuroimaging studies [18–22] mainly focused on the conversion from MCI to AD [23]. Van Loenhoud et al. analyzed the differences between predicted brain damage on neuroimaging and cognitive testing. They found that less brain damage than expected was a predictor of lower conversion from normal to MCI or AD, but they did not provide prediction at the individual level [24]. Liu et al. treated the transition as a regression problem, which did not use longitudinal information [25]. In addition, longitudinal MRI has also been used to predict brain age, highlighting the accelerated biological aging in individuals who develop AD dementia [20]. While significant advancements have been achieved during recent years, there remain substantial challenges on effectively distinguishing the cognitive decline linked to AD from that associated with normal aging [19–22].

Deep learning-based approaches such as the convolutional neural network (CNN) have recently become popular for brain imaging data analysis, including image classification, abnormality detection, and even early diagnosis of various diseases [26]. These approaches have two advantages: 1) they can process large amounts of data quickly and accurately, and 2) they can detect patterns or features in complex data that are invisible to the human eye. MRI is a medical imaging technique used to diagnose and monitor various medical conditions. Many studies have demonstrated the utility of CNNs in accurately classifying the different stages of dementia based on MRI data [27–30]. Despite the advancements in deep learning, large-scale MRI studies have a noticeable research gap. The field of neuroscience research often faces the challenge of small sample sizes, which can result in overfitting and reduced model generalizabil-

ity. To fill in this gap, we propose to transform the supervised prediction problems, such as AD versus cognitively normal (CN), into a self-supervised contrastive learning problem on a longitudinal follow-up for each subject via extracting changes in both neuroimaging and cognitive assessment scores over time, which might have a better power to solve the current limitations in the field.

In this study, we combined neuroimaging, clinical, and genotyping data to create a comprehensive deep learning-based method for disentangling accelerated cognitive decline from the normal aging process and exploring its underlying genetic basis. We hypothesized that the well-fitted neuroimaging model trained on the population at large can be applied to all subjects to capture the normal cognitive decline due to normal aging. For a comprehensive understanding, we compared our model's performance with two previous published studies that employed Recursive neural network (RNN) [31] and rank-based CNN [32] models on longitudinal data. Our approach, which tested on Alzheimer's Disease Neuroimaging Initiative (ADNI) cohorts and outperformed cognitive decline alone, revealed new loci and genes not previously identified in AD studies [6, 33]. Our work presents three major contributions. First, we created data pairs involving T1-weighted MRI (T1w MRI) and corresponding ADAS-Cog13 neuropsychological assessment results for all possible combinations of time points within the set. These data pairs were then trained with a dual-loss Siamese ResNet model to assess whether a pair of MRI images and cognitive score alterations exceeded a certain normal aging threshold. We applied the pre-trained model to predict aging-related cognitive decline for the population at large. By accounting for the confounding factor of normal aging, this model enhances the statistical power of subsequent genome-wide association studies (GWAS) focused on accelerated cognitive decline. Secondly, we adopted metric learning and multitask learning by combining supervised learning and self-supervised contrastive learning tasks on a continuous severity scale of MRIs and cognitive assessment. Our model could use unlabeled data to learn both similarities and dissimilates between pairs of MRI images, resulting in a robust vector representation of an MRI image that is not dependent on the ground-truth label. Therefore, the learned image representations are robust to label bias and are more generalizable. Tests conducted on ADNI cohorts encompassing CN, MCI, and AD individuals demonstrated that our model outperforms ADAS-Cog13 items, as evidenced by reduced standard error and dispersion measures in the cognitive decline rate. Lastly, our GWAS and subsequent post-GWAS analyses successfully identified novel loci and genes that had remained undiscovered in previous AD GWAS studies.

METHODS

Alzheimer's Disease Neuroimaging Initiative (ADNI) data

Data used in the preparation of this article were obtained from the ADNI database (http://adni.loni.usc.edu). The 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 MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see http://www.adni-info.org.

In this study, we used the ADNI database (ADNI 1, GO, 2, 3) to build the imaging-cognitive score model. The longitudinal analysis of T1w MRI data was used to provide brain structural information of both gray and white matter to track and evaluate brain structural change along the time axis as the disease progresses. We paired T1w MRI images from 1,313 participants with their cognitive score tests assessed from 2003 to 2019; the age of the participants covered a wide spectrum ranging from 55 to 91 years old. The 1,313 participants were categorized as CN, MCI, or AD based on their cognitive status at the baseline screening for training the deep learning model. ADNI demographic information used for MRI image training and prediction was summarized in Supplementary Table 1.

ADAS-Cog assessment

We used the ADAS-Cog13 items scores from ADNI clinical data for the cognitive score assessment. ADAS-Cog13 was developed to be used as an index of global cognition in disease progression assessment. ADAS-Cog13 includes 13 items assessing cognitive function [34]. The tasks are related to memory, language, praxis, orientation, number cancellation, and delayed free recall, with a total score of 85 points, with higher scores denoting worse performance.

Image preprocessing

We curated longitudinal 6,711 T1w MRI images from 1,313 participants with paired ADAS-Cog 13 assessments and processed them using Clinica [35] to the Brain Imaging Data Structure (BIDS) format [36]. MRI images were first processed using a nonparametric nonuniform intensity normalization (N3) algorithm [37] to correct the non-uniform intensity. After correction, skull-stripping was performed by PARIETAL [38], followed by registering MRI images to a common template (Montreal Neurological Institute 152) using Freesufer 6.0.1 [39] and removing voxels outside the brain region. All images were prepared in $128 \times 160 \times 128$ resolution and $1.0\,\mathrm{mm}^3$ voxel size.

Deep learning pipeline

Experimental design

Accelerated cognitive decline was defined as having a steeper slope in the cognitive assessment than normal aging. To calculate the cognitive decline slope, we selected subjects with more than two visits of paired ADAS-Cog13 assessments and T1w MRI scans. Specifically, the ADAS-Cog13 subtest was linked to one T1w MRI scan if it was tested within 45 days of the T1w MRI assessment. We only included individuals with their diagnosis status either unchanged or having forward transitions among data collection points. To capture a stable cognitive alteration trend, we implemented a time span ranging from a minimum of 6 months to a maximum of 24 months between data collection points to form pairs and excluded subjects with no more than 2 data points (Supplementary Figure 1). By applying a skip connection between longitudinal data points, we were able to curate 9,680 data pairs from 1,313 subjects. We used a linear relationship to represent the proximity of the cognitive changes along the progression of AD [40–43]. Therefore, the median cognitive decline slope ($\frac{ADAS-Cog13t_i - ADAS-Cog13t_{i-1}}{t_i - t_{i-1}}$, Fig. 1A) across visits was used as the outcome of the following GWAS.

Deep learning architecture

We illustrated our overall deep learning framework, which employs dual-loss Siamese ResNet, in Fig. 1A. First, the multitask neural network was trained to simultaneously perform two tasks: predicting the actual cognitive score at the first time point of the data pair (regression task) and distinguishing a pair of images belonging to the same/different classes (contrastive learning task). Second, the two tasks share a common backbone neural network structure, which has a similar structure to the Siamese network [44]. The output of the network has two pre-

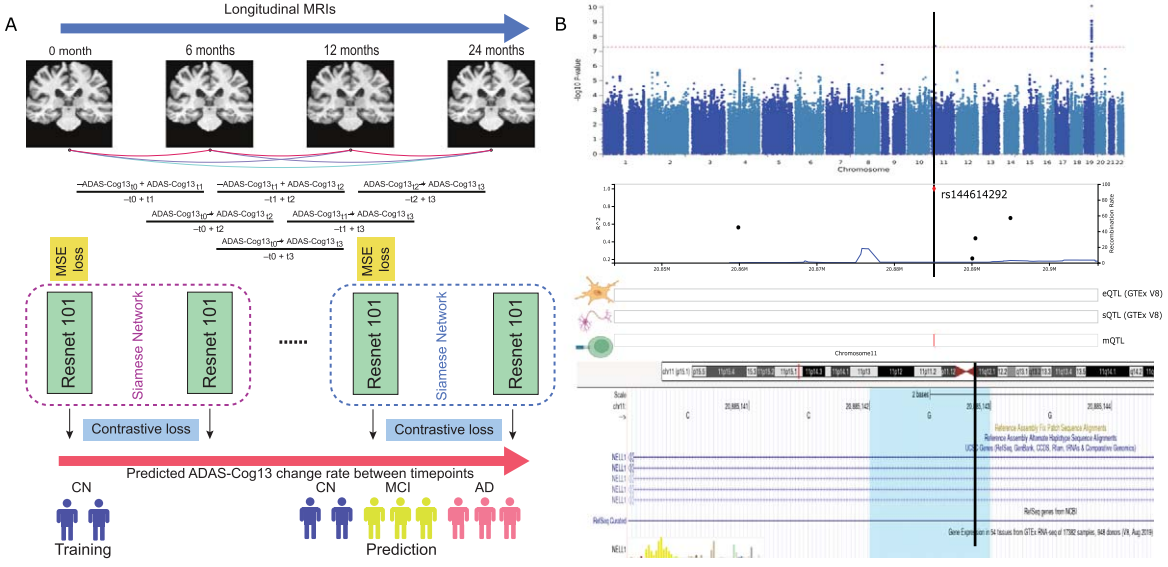

Fig. 1. Overview of this work. A) Deep learning of accelerated cognitive decline. B) GWAS analysis reveals one novel locus related to accelerated cognitive decline.

diction heads with a Multilayer Perceptron Network structure to perform the two tasks. The model takes paired two separate images from two time points as input, feeds them into the shared subnetworks, and joins the two output embedding vectors to feed into separate task-specific layers [45].

To extract features from MRI data, we used 3D ResNet-101 [46] as subnetworks with shared weights using 3D kernels instead of original 2D kernels. We first introduced mean square error (MSE) loss to counteract baseline differences between pairs by ensuring the predicted ADAS-Cog13 scores are closely aligned with true target values for the first point of each time pair. We skipped the final fully connected layer and used the high-dimensional vector output to calculate the Euclidean distance between subnetworks. While using the paired image input X_1 and X_2, we calculate the Euclidean distance between the subnetwork output vectors $Gw(X_1)$ and $Gw(X_2)$ as $Dw(X_1, X_2) = \|Gw(X_1) - Gw(X_2)\|_2$. Then, we introduced contrastive loss as $L = (1 - Y)\,Dw^2 + (Y)\,\{max(0, m - Dw)\}^2$, where Y is the actual label of a pair of MRI images ($Y = 0$ if belonging to the same class, i.e., no significant change on cognitive score; $Y = 1$ if belonging to different classes, i.e., significant change on cognitive assessment scores). We chose an absolute value of 5 as the empirical threshold (>∼95 percentile as the outliers) to label our data. The variable m is a hyperparameter denoting the minimum Euclidean distance (ED) a pair of different-class images should have. In the training analysis, 1,959 data pairs from 289 CN subjects were used. We employed the Adam optimizer [47] and a mini-batch size of 4 to train the model for 200 epochs with an initial learning rate of 10^{-4} and a step-based learning rate scheduler with a decay rate $\gamma = 0.1$ for every 10 epochs on a Nvidia-A100 GPU. In validation, 946 data pairs from 125 CN subjects were used to test the performance of the best model, with minimum validation loss summation of MSE and contrastive loss. Lastly, we predicted all the cognitive decline slopes as changes of ADAS-Cog13 scores divided by time for each subject at each pair as the learned cognitive effect from neuroimaging. Finally, the median of the predicted cognitive decline slope between visits for each subject was used as the covariate of the GWAS.

Performance evaluation

The normalized predicted error, defined as the difference between predicted and actual ADAS-Cog13 scores divided by time), was used to measure the model performance among dual-loss Siamese ResNet networks with different depths (101, 152, 200) of 3D ResNet subnetwork structures. To verify the stability of our framework, we further compared their GWAS analysis results, including the lead SNPs and Manhattan plots, respectively. We selected our best dual-loss Siamese ResNet model using 3D ResNet-101 as subnetworks and compared the normalized predicted error of model performance with other two existing deep learning methods (Comparison 1: Ranking convolutional neural network [46], Comparison 2: Recursive neural network [31]).

Quality control and GWAS

Imputation process

We obtained the ADNI raw genotype data from the ADNI [48], including three batches of study (ADNI1 757, ADNI2GO 793, and ADNI3 327 individuals). We followed the procedure of previous work [49]. Briefly, we first converted all the SNPs to human reference (GRCh37) using liftover [50]. Before imputation, we performed the standard variants checking procedure to correct abnormal SNPs using the tools developed by the McCarthy group [51]. Then, we submitted all the pre-checked genotype data to the Michigan Imputation Server [52], using the 1000g-phase-3-v5 European ancestry reference panel, respectively. Next, we combined these three cohorts and filtered out those imputed variants with imputation quality <0.6, the remaining 12,144,319 variants in total shared between three batches of ADNI cohorts.

Quality control (QC) analysis

We applied KING v.2.2 [53] to remove individuals estimated to be closer than second-degree relatives with a kinship coefficient >0.0884, which kept 1,858 out of 1,877 total individuals. ANNOtate VARiation (ANNOVAR) [54] was used to annotate the rsid of each SNP from dbSNP151. Next, we used bcftools [55] and vcftools [56] to replace the ID column of the vcf file. Next, we adopted plink1.9 [57] to conduct the standard QC procedures, including SNP missing rate >0.02, minor allele frequency >0.01, and Hardy-Weinberg Equilibrium $>10^{-6}$. Overall, we obtained 8,261,400 variants for GWAS analysis for 1847 individuals.

European ancestry (EA) cohort population

The ADNI cohorts are largely of European ancestry (EA). Therefore, we extracted EA subjects by projecting them into the 1000 Genomes Project individuals with different ethnic backgrounds. First, we pruned the SNPs using the command '–indep-pairwise 50 5 0.2' from plink, which greedily pruned 5 pairs of variants in the 50 kb window with a squared correlation greater than 0.2 until no such pairs remained from the window. We downloaded the genotype information of 629 individuals from the 1000 Genomes Project ftp [58]. We selected the previous SNPs after pruning and merging these 629 individuals with our 1858 ADNI participants. We conducted a multidimensional scaling (MDS) analysis to identify the population stratification. We excluded the outliers from EA (Supplementary Figure 2A). After overlapping with samples with longitudinal MRI data (1290), 1064 individuals with EA were retained for downstream GWAS analysis (Supplementary Table 2).

GWAS for cognitive decline slope

In this work, we explored the genetic variants that contributed to the accelerated cognitive decline slope. We applied two linear regression models to conduct the GWAS analysis on the ADNI cognitive decline slope and accelerated cognitive decline slope.

Model 1: median Cog decline slope $\sim$ genotype + median predicted aging-related Cog decline slope + PCs + sex + median measured age.
Model 2: median Cog decline slope $\sim$ genotype + PCs + sex + median measured age, where PCs are the top 10 principal components (PCs) from the multidimensional scaling (MDS) analysis of 1064 genotype data with previous pruned SNPs. Sex information is adopted from the ADNI demographic annotation. The median predicted aging-related Cog decline slope is derived from the pre-trained model, as mentioned in the deep learning architecture session. To increase the power and deflated type I error in non-normally distributed quantitative traits, we applied the inverse normal transformation to normalize the measured age, cognitive decline slopes and predicted the slope using r package RNOmni [59].

Lead SNPs, QTL traits, and colocalization analysis

We defined the lead SNPs with nominal significance ($p < 10^{-5}$). We pruned their nearby SNPs with LD $r^2 \geq 0.6$. Then, the remaining SNPs with LD $r^2 \geq 0.1$ were pruned to define the independent lead SNPs. These independent loci were combined if they were separated by less than 250 kb.

To understand the potential functions of these variants among different tissues and cell types, we scanned the top three SNPs with $r^2 > 0.4$ of the lead SNPs of interest [chromosome (chr)4-rs4694308 and chr11-rs144614292] among thousands of quantitative trait loci (QTL) resources curated in QTLbase v2.2 [60]. We selected the potential QTL traits associated with the SNPs of interest that have significant signals within the high LD region of SNPs of interest. To understand the single-cell QTL in the brain, we adopted the latest brain cell eQTL dataset [61] as well. Colocalization analysis was performed using Bayesian Coloc [62], which aims to identify a genetic variant that has shared causality between expression and GWAS trait. The Coloc script was extracted from the original Coloc package [63] and revised as our previous work [64, 65]. The posterior probability of H4 > 0.5 was defined as nominal significance.

Phenotype-wide association studies (PheWAS)

To explore the biological insight of the identified statistically significant variants, we assessed the PheWeb version 1.3.15 [66] to query their impacts in $\sim$1400 Phenome-wide association studies (PheWAS) conducted in the UK Biobank cohort. Considering the potential correlation between SNPs within the high linkage disequilibrium region, we checked the top three SNPs with $r^2 > 0.4$ of the lead SNPs of interest (chr4-rs4694308 and chr11-rs144614292).

Gene-level p-value and over-representative analysis

Gene-level *p*-value was precalculated by MAGMA [67] (incorporated in FUMA platform) with a 50 kb SNP window surrounding each gene. Then, we performed the gene-set analysis implemented in Functional Mapping and Annotation of GWAS [68–70], which utilizes a linear regression to test if the conditional (such as gene length and gene correlation) mean association with the cognitive function decline phenotype of genes in curated gene sets is greater than that of genes not in the gene set. The cognitive function gene sets were defined by the 52 genes mapped from all the lead SNPs within 50kb in the FUMA platform (Supplementary Table 3). In total, 15,487 gene sets [C2 and Gene Ontology (GO) terms] from

the Molecular Signatures Database (MSigDB) [71] were used to test the functional over-representation.

Tissue and cell-type specific enrichment analysis

We adopted the MAGMA tissue-specificity test deployed in FUMA, which performs a linear regression, to test if the cognitive function decline phenotype of genes is more expressed in a specific tissue when compared to other tissue types for 53 tissues from GTEx V8 [72].

To understand the cell-type-specificity of the target GWAS genes, we adopted our in-house online tool, Web-based Cell-type Specific Enrichment Analysis (WebCSEA) [73]. This platform utilizes our previous deTS algorithm [74] to calculate the raw *p*-value across 1,355 tissue-cell types curated from the large consortium datasets [75]. A permutation-based test was applied to overcome the potential bias due to the different lengths of signature and type I errors. Specifically, we calculated the permutation *p*-value by ranking the queried raw *p*-value over more than the *p*-values of 20,000 gene lists from GWAS and a rare-variants association study of human complex traits and disease pre-curated in WebCSEA. We adopted the 52 genes mapped from all the lead SNPs within 50kb in FUMA platform to WebCSEA. The suggestive significance was set to 0.001. In addition, we check the tissue and cell type implications of all lead SNPs using our in-house method DeepFun [76, 77], which utilizes the convolutional neural network framework to predict the SNP Activity Difference (SAD) on ∼8,000 chromatin profiles of 225 tissues or cell types from Encyclopedia of DNA Elements (ENCODE) and Roadmap projects.

Polygenic risk score (PRS) analysis

LDpred2 [78] was used to calculate the polygenic risk score (PRS). We adopted the summary statistics from the meta-analysis of AD GWAS by Wightman et al. [6]. This meta-analysis excluded the proxy cases from UK Biobank and 23andMe subjects, which included 39,918 cases and 358,140 controls. We utilized HAPMAP3 variants identified in both GWAS summary statistics and the genotype data of the ADNI cohort. Based on matched SNPs, LDpred2(-grid) was used to calculate the candidate PRS for each individual in the ADNI cohort with each hyperparameter combination. We calculated the same for different *p*-value thresholds (1, 0.5, 0.3, 0.1, 0.05). The PRS was generated by selecting the hyperparameter combination that achieves the highest area under the curve (AUC) when using the AD diagnosis as the reference group.

Two-sample Mendelian randomization analysis

Two-sample Mendelian randomization (2SMR) is a statistical method leveraging independent GWAS summary statistics to evaluate causality between an exposure and an outcome using genetic variants as instrumental variables [79]. Here, we conducted a 2SMR analysis to assess the causality of the association between cg07126637 and cognition variation using the R package 2SMR [79, 80]. We first obtained all the methylation qualitative trait locus (mQTL) within the region of cg07126637 from one previous genome-wide mQTL study [81]. Considering that mQTLs may be associated with cg07126637 due to linkage disequilibrium (LD) patterns, we performed LD clumping on mQTLs to remove all SNPs present in the 1000 Genome European population with $r^2 > 0.1$ and within 10 kb of the top SNPs. We then extracted and harmonized matched SNPs from our GWAS summary statistics. Finally, we performed 2SMR on the harmonized data using built-in methods in the package, including inverse-variance weighted and Egger, among others.

Genetic correlation analysis

We calculated the liability-based heritability and the magnitude of genetic correlation between AD and other cognitive function-related phenotypes (Supplementary Table 4) using the LD score regression model [82]. Pre-estimated LD scores were obtained from the 1,000 Genomes Project European reference population, and then we calculated the genetic correlation employing HapMap3 SNPs only with LD reference panel SNPs to minimize potential bias due to differences in LD structure.

RESULTS

Measuring the accelerated cognitive decline from longitudinal neuroimaging data

To disentangle the impact of cognitive decline due to the normal aging process from accelerated aging, we developed a deep learning framework that employs dual-loss Siamese ResNet. This framework enables better prediction of longitudinal cognitive score decline of individuals by extracting the imaging

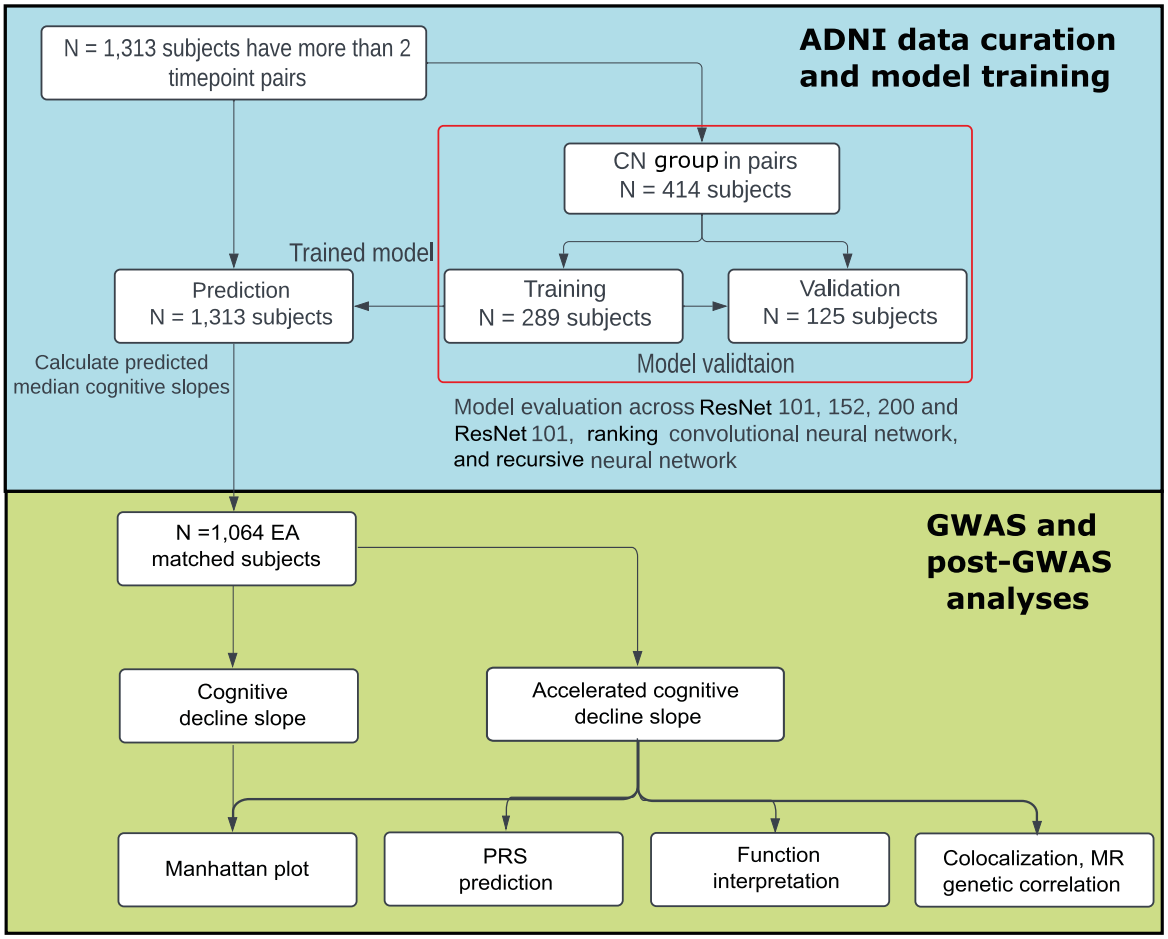

Fig. 2. Workflow of the deep learning model design of accelerated cognitive decline prediction using ADNI data and GWAS/post-GWAS analysis.

features and leveraging temporal correlations with paired T1w MRIs. As illustrated in Fig. 1A and Fig. 2, we obtained the matched brain imaging, clinical data (cognitive assessment, ADAS-Cog13), and genotype data. The longitudinal ADAS-Cog13 scores for all 1,313 subjects were considered. We could observe a clear separation among CN, MCI, and AD (Fig. 3A). We defined the cognitive decline slope between time points t_i and t_j as $(\Delta CS_{t_i,t_j})$ divided by $(t_i - t_j)$ (Fig. 1A, Fig. 3B).

For dual-loss Siamese ResNet, we used 3D ResNet-101 as the subnetworks backbone to extract the paired MRIs (X_{t_i}, X_{t_j}) data into embedding vector $Gw(X_1)$ and $Gw(X_2)$. Their difference was defined as the Euclidean distance $Dw(X_1, X_2) = \|Gw(X_1) - Gw(X_2)\|_2$. We leveraged the dual loss design to further capture the similarity/difference between paired MRIs (X_{t_i}, X_{t_j}). We trained and validated our model on 414 CN individuals in a 70/30 splitting ratio and predicted the cognitive assessment in 1,313 individuals. Model performance was evaluated using the normalized predicted error (NPE, difference of predicted and actual ADAS-Cog13 divided by time) of predicted cognitive decline slopes in the validation cohort (946 pairs from 125 CN individuals). The accelerated cognitive decline slope (Fig. 3C) was calculated as the residual of cognitive decline slope (Fig. 3B) by predicted aging-related cognitive decline slope using linear regression [Model 1], see "Methods"). In Fig. 3D, we conducted pairwise Wilcox tests for three clinical diagnoses. Except for CN versus MCI, which were not significant ($p = 0.058$), all other comparison groups showed a significant difference.

In addition, the variance of the estimated cognitive decline related to normal aging increased along with the clinical diagnosis, suggesting a larger variation in the CN group compared to the MCI and AD

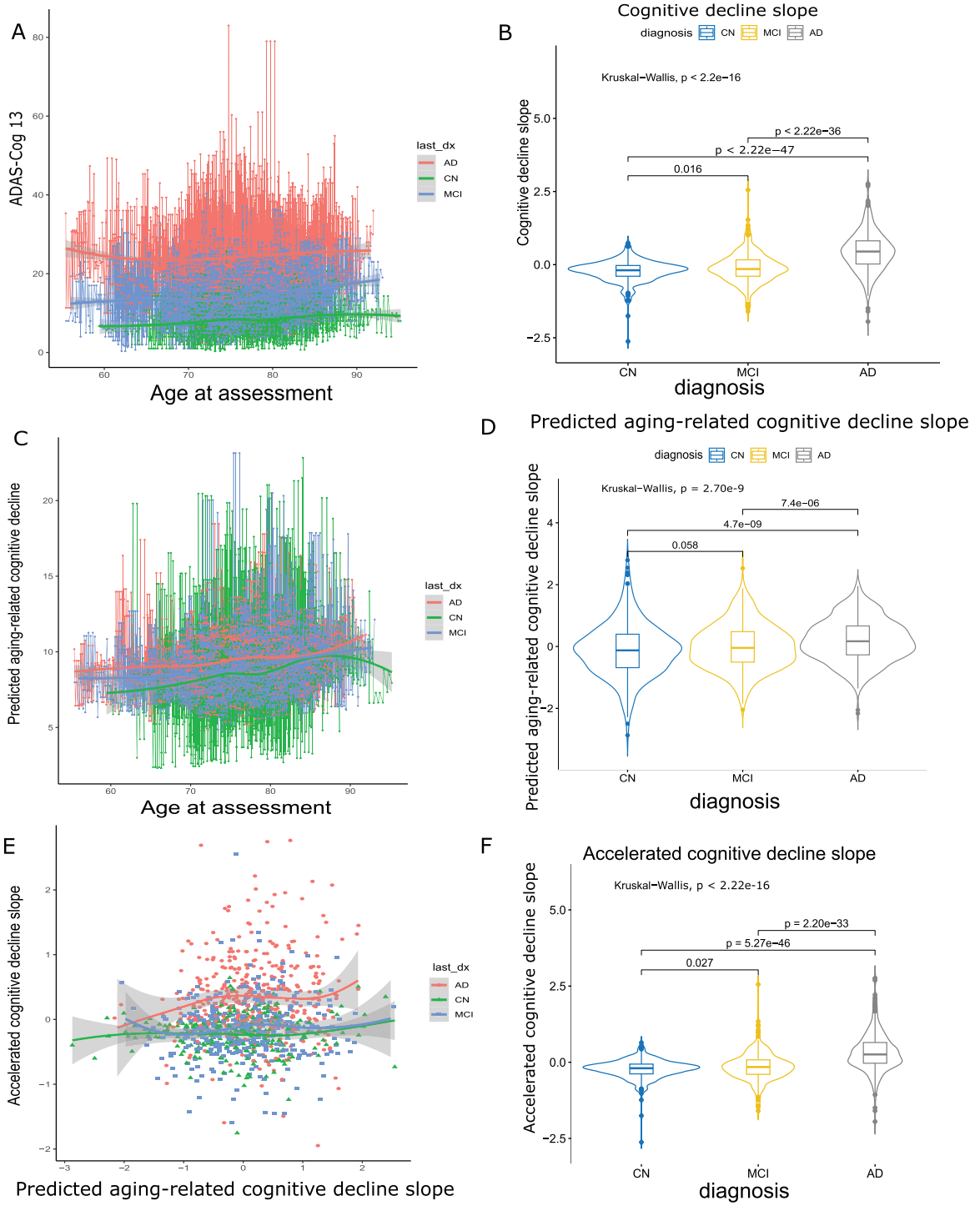

Fig. 3. Cognitive decline and cognitive decline slope visualization. A) A longitudinal ADAS-Cog13 assessment for each individual by age at measurement stratified by clinical diagnosis. Each dot represents one measure. B) Raw cognitive decline slope distribution by clinical diagnosis. C) Predicted longitudinal aging-related cognitive decline by age at measurement stratified by clinical diagnosis. Each dot represents one measure. D) Predicted aging-related cognitive decline slope distribution by clinical diagnosis. E) Association between predicted aging-related cognitive decline slope and accelerated cognitive decline slope stratified by clinical diagnosis. F) Accelerated cognitive decline slope distribution by clinical diagnosis. last_dx, last clinical diagnosis.

groups. In the AD group, we uncovered a positive correlation trend between the aging-related cognitive decline slope and the accelerated cognitive decline slope (Fig. 3E), while no such tendency was observed in the CN or MCI groups. This observation indicates a distinctive effect of the accelerated cognitive decline slope within the AD group. The distribution of the accelerated cognitive decline slopes (Fig. 3F) was also verified in the observations shown in Fig. 3E. In contrast to the original cognitive decline slope across the clinical diagnoses (Fig. 3B), the significance level of accelerated cognitive decline (Fig. 3F) is slightly smaller, indicating that the cognitive decline linked to the accelerated cognitive decline slope exhibits a closer magnitude in comparison to the cognitive decline slope across the diagnosis groups. The distribution differences (p values) among diagnosis groups are considerably more prominent in both the cognitive decline slope (Fig. 3B) and the accelerated cognitive decline slope (Fig. 3F) than in the predicted aging-related cognitive decline slope (Fig. 3D). This suggests that cognitive decline associated with normal aging exhibits a smaller magnitude when contrasted with the cognitive decline linked to AD. We further compared our model (NPE = –0.29, $\sigma = 0.0040$) with two different deep-learning model designs in Supplementary Figure 3A (Comparison 1: ranking convolutional neural network [CNN] [33] (NPE = 0.042, $\sigma = 0.0053$) and Comparison 2: recursive neural network [RNN] [32] (NPE = –0.34, $\sigma = 0.0040$)) and showed that our model has significantly better performance (more constrained error dispersion). Lastly, dual-loss Siamese ResNet with different depths of 3D ResNet Subnetwork [101, 152, 200] shows similar performance in Supplementary Figure 3B.

One novel locus identified by GWAS of accelerated cognitive decline

We formulated two different models to capture the genetic basis that contributes to the cognitive decline slope: 1) an accelerated cognitive decline slope and 2) the original cognitive decline slope. We followed the illustration in Fig. 1B to conduct a comprehensive post-GWAS analysis to interpret the genetic factors associated with accelerated cognitive decline. As shown in Fig. 4A, the following GWAS for accelerated cognitive decline slope identified two genome-wide significant loci (chr11 rs144614292:G>T $p = 3.73 \times 10^{-8}$ and chr19 rs429358 in *APOE* locus). The rs144614292, with a minor allele frequency of 0.05 in the EA population, is an intronic variant of the *NELL1* gene, which encodes for the teneurin-2 protein and plays a role in synaptogenesis, neurite outgrowth, axon guidance, and neuronal connectivity [83]. In total, we observed 21 loci with suggestive significance ($p < 10^{-5}$) (Table 1). As shown in Fig. 4B, only the chr19 *APOE* locus was identified in the original cognitive decline GWAS. We further checked the PRS of AD for these 1,064 individuals using the weight from one previous AD GWAS summary statistics [6]. We identified that the individual PRS is positively correlated with the severity of the clinical diagnosis and is significantly different between diagnostic categories (Fig. 4C). Lastly, we identified that the AD PRS is positively correlated with the normalized cognitive decline slope (Fig. 4D).

Gene-based value highlights one significant region

We conducted the gene-level p-value using MAGMA [68]. Except for the known *APOE* region genes, we identified two significant genes, MUC7 (1.12×10^{-6}) and PROL1 (1.42×10^{-6}), after Bonferroni correction ($0.05/19{,}171 = 2.61 \times 10^{-6}$) in the chr4 lead region locus (rs4694308C>T) (Supplementary Figure 4A). This high LD region ($r^2 > 0.6$) expands about 0.14 million bps (chr4 : 71.26M-chr4 : 71.40M) and contains five genes (*SMR3A, SMR3B, PROL1/OPRPN, MUC7, AMTN*) directly overlapped with high LD region (Supplementary Figure 5). Another two genes, (*CABS1* and *AMBN*), are within 100 kb of this locus, making it challenging to map the risk variant to the corresponding genes. On the other hand, the original cognitive slope GWAS did not identify other significant genes, except the chr19 *APOE* region (Supplementary Figure 4B).

Colocalization and Mendelian randomization

To verify the novel SNPs and genes findings, we conducted colocalization for two major regions of interest: the gene-level significant locus chr4 rs4694308 C>T (Fig. 5C) and the novel genome-wide significant locus chr11 rs144614292 G>T (Fig. 5D), respectively. We collected one single-cell brain-related eQTL dataset [61] and another blood QTL dataset [81] that is highlighted in the QTLbase [60]. We adopted the colocalization method Coloc and QTL data resource (Fig. 5A and Table 2). The only significant PP H4 cg07126637 (0.68) is visu-

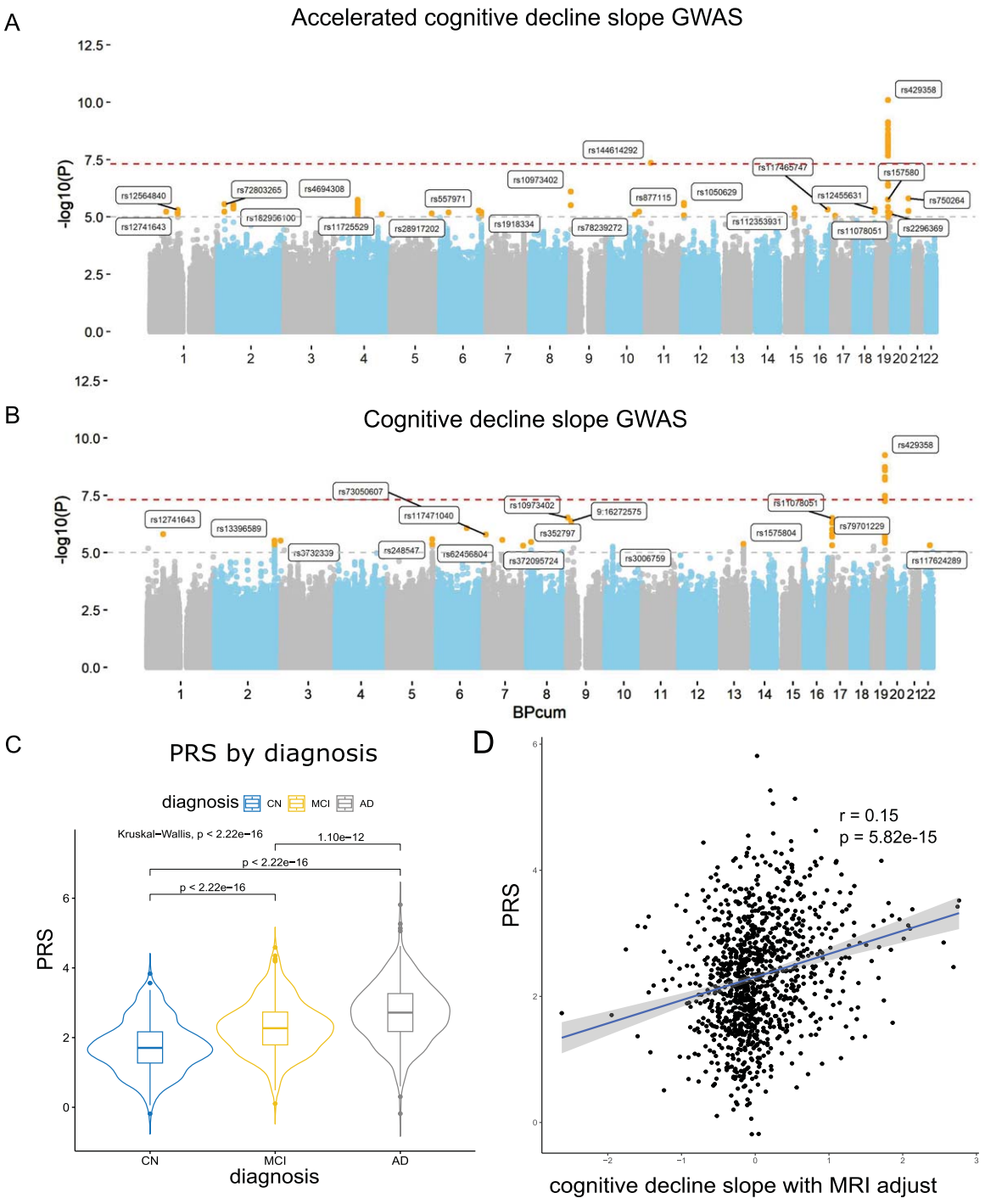

Fig. 4. GWAS and PRS analyses. A) Manhattan plot for accelerated cognitive decline slope. B) Manhattan plot for original cognitive decline slope. The lead SNP in each locus was highlighted. C) AD clinical diagnosis by AD PRS distribution. CN, cognitively normal; MCI, mild cognitive impairment. D) Correlation between AD PRS distribution and accelerated cognitive decline slope.

alized in the 2SMR analysis (Fig. 5B). A total of 12 SNPs was included in the analysis after LD clumping and harmonization procedures. The results showed that cg07126637 CpG site was significantly associated with cognitive decline slope using the inverse variance weighted method (beta = –0.327,

Table 1
List of lead SNPs with $p < 1 \times 10^{-5}$

Locus	rsID	chr	pos	ref	alt	beta	se	p
1	rs12741643	1	60752854	A	G	0.19	0.042	5.90×10^{-6}
2	rs12564840	1	102589297	A	G	0.27	0.059	4.91×10^{-6}
3	rs72803265	2	24724907	C	G	0.49	0.11	2.79×10^{-6}
4	rs182956100	2	56783212	C	T	0.24	0.05	2.84×10^{-6}
5	**rs4694308**	**4**	**71284706**	**C**	**T**	**0.27**	**0.057**	$\mathbf{1.83 \times 10^{-6}}$
6	rs11725529	4	159211118	A	G	–0.72	0.16	7.65×10^{-6}
7	rs28917202	5	151128193	C	T	–0.68	0.15	7.00×10^{-6}
8	rs557971	6	143030052	A	T	–0.61	0.13	5.37×10^{-6}
9	rs1918334	6	152925572	G	T	0.22	0.049	6.13×10^{-6}
10	rs10973402	9	3763389	A	G	–0.22	0.045	7.93×10^{-7}
11	rs78239272	10	98460748	C	G	0.90	0.20	8.02×10^{-6}
12	rs877115	10	112104806	A	G	–0.21	0.046	6.02×10^{-6}
13	**rs144614292**	**11**	**20885143**	**G**	**T**	**–0.67**	**0.12**	$\mathbf{4.38 \times 10^{-8}}$
14	rs1050629	12	6857289	A	G	0.65	0.14	2.49×10^{-6}
15	rs112353931	15	55868481	C	T	0.76	0.16	4.10×10^{-6}
16	rs117465747	16	73587857	G	T	0.62	0.14	4.81×10^{-6}
17	rs11078051	17	12057796	C	T	–0.22	0.050	8.72×10^{-6}
18	rs12455631	18	75082700	A	G	–0.67	0.15	4.68×10^{-6}
19	rs157580	19	45395266	A	G	–0.21	0.045	1.77×10^{-6}
19	**rs429358**	**19**	**45411941**	**C**	**T**	**–0.30**	**0.046**	$\mathbf{8.02 \times 10^{-11}}$
20	rs2296369	19	55224544	A	C	–0.30	0.067	7.20×10^{-6}
21	rs750264	20	61993767	C	G	–0.36	0.075	1.55×10^{-6}

chr, chromosome; pos, position; ref, reference allele; alt, alternative allele; se, standard error. Loci of interest were in bold.

$p = 5.49 \times 10^{-5}$). There was no horizontal pleiotropy according to the MR Egger regression test ($p = 0.81$). Among the single SNP analysis, rs777390 (GWAS $p = 2.89 \times 10^{-4}$) had the most significant results with $p = 2.76 \times 10^{-4}$.

Functional interpretation of genetic factors associated with accelerated cognitive decline

To assess how these genetic factors manifest their effect on tissue and cell types, we applied FUMA MAGMA tissue-specificity test across 53 tissues from GTEx V8. We identified the liver, skin, esophagus mucosa, prostate, brain, and spinal cord cervical as the top five tissues. However, none of them were significant (Fig. 6A). The WebCSEA analysis suggests that thymocyte (combined $p = 4.93 \times 10^{-5}$), stromal cell (combined $p = 3.74 \times 10^{-4}$), and microglia (combined $p = 1.64 \times 10^{-3}$) are the top three cell types related to cognitive decline (Fig. 6B). We applied 21 independent lead SNPs to the DeepFun Web service (Fig. 6C). The chr4 lead SNP (rs4694308 C>T) does not find SNP Activity Difference (SAD). In contrast, chr11 lead SNP (rs144614292 G>T) was found to have SAD signals in the brain, particularly the frontal cortex. Universal SAD alterations could be observed in chr19 lead SNP (rs429358 $T > C$), suggesting the regulatory effect could impact most tissue and cell types. The functional over-representative analysis of 52 genes mapped from the lead SNPs in MSigDB (C2 and GO terms) highlighted lipid metabolism and immune response functions, aligned with our previous tissue (liver) and cell-type enrichment findings (thymocyte and microglia) (Supplementary Figure 6). Lastly, no PheWAS conducted within the UK Biobank cohort revealed significant associations ($p < 0.05/1419$ phenotypes) with the lead SNPs of interest (chr4-rs4694308 and chr11-rs144614292), suggesting no known associations between the two loci with recorded phenotypes (Supplementary Figure 7).

Genetic correlation suggests cognitive decline is positively associated with AD

We did a pairwise genetic correlation comparison between the following traits: AD, accelerated cognitive decline slope, original cognitive decline, and educational attainment (Supplementary Tables 4 and 5, Supplementary Figure 8). As expected, AD was negatively correlated with educational attainment (genetic correlation (r_g) = –012, $p = 0.020$), the only significant r_g identified. Positive, although not significant, correlations were observed in AD versus

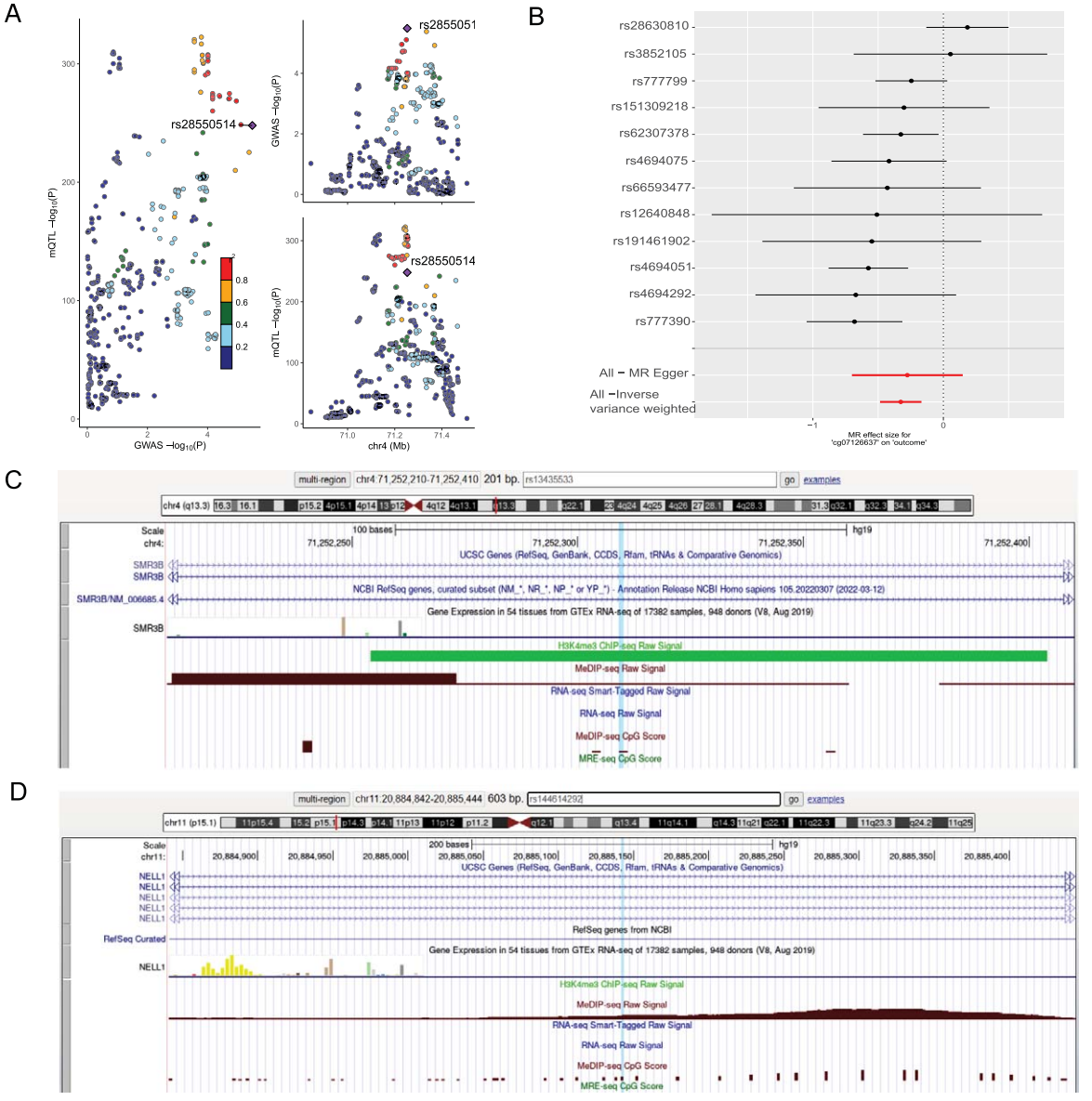

Fig. 5. A) LocusZoom plot at chr4 rs4694308C > T locus for exposure mQTL associated with cg07126637 and outcome SNPs associated with accelerated cognitive decline slope. B) Mendelian randomization analysis for 12 independent SNPs from 500kb up- and downstream of rs4694308 between mQTL (cg07126637) and SNPs associated with accelerated cognitive decline slope. All MR Egger results and Inverse variance weighted tests were highlighted. C) UCSC Genome Browser view of chr4 rs4694308C > T locus indicated that the merged lead SNP of rs13435533 is impacted by DNA methylation and H3K4me3 signals. D) UCSC Genome Browser view of chr11 rs144614292G > T locus indicated that the lead SNP rs144614292 is impacted by DNA methylation signal.

Table 2
Summary of colocalization analysis results at the chr4 and chr11 loci

Quantitative traits	Locus	Tissue/cell type QTL	QTL ref.	Coloc PP.H4
cg07126637 chr4 : 71248757 (intron region of SMR3B)	rs4694308 (chr4)	Blood mQTL	34493871	0.68*
cg03970609 chr4 : 71337664 (intron region of MUC7)	rs4694308 (chr4)	Blood mQTL	34493871	0.14
NELL1	rs144614292 (chr11)	Excitatory Neurons eQTL	35915177	0.086

*Indicates nominal significance of posterior possibility (PP H4 > 0.5) in Coloc analysis.

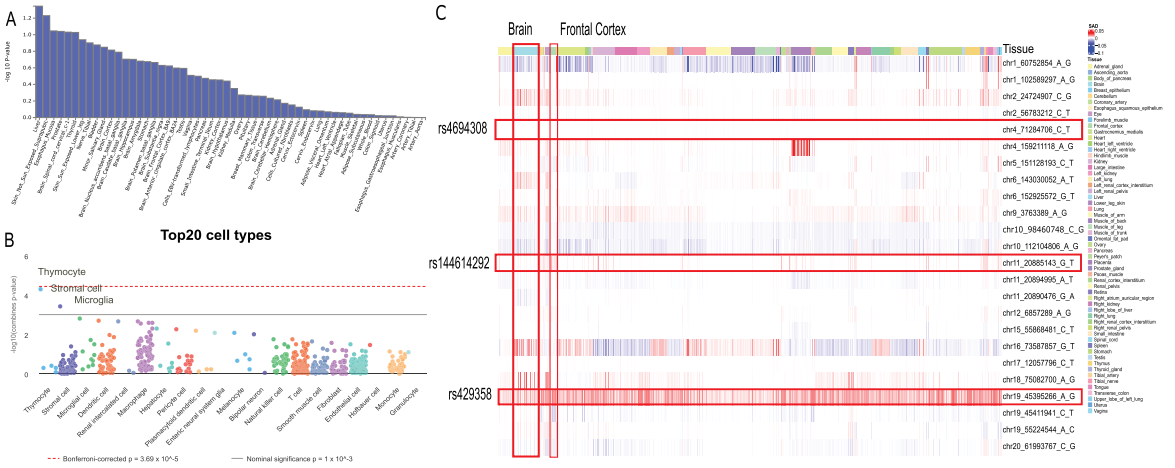

Fig. 6. Tissue and cell-type specificity of cognitive decline factors. A) Tissue-specificity. B) Cell-type specificity. The top three general cell types were highlighted. C) DeepFun shows the SNP Activity Difference (SAD) scores across tissue and cell types. The lead SNPs of interest and brain tissues were highlighted.

accelerated cognitive decline slope ($r_g = 0.1, p = 0.65$) and Wightman AD versus original cognitive decline ($r_g = 0.50$, $p = 0.20$). The original cognitive decline versus educational attainment [84] has a light positive correlation ($r_g = 0.04$, = 0.65), while the accelerated cognitive decline versus education attainment has a larger positive correlation ($r_g = 0.10$, $p = 0.5$). However, again, they were not statistically significant.

DISCUSSION

We developed a novel deep-learning-based approach, leveraging dual-loss Siamese ResNet to learn the normal aging-related cognitive decline slope and identified the underlying genetic risks for accelerated cognitive decline. Besides the well-known *APOE* region, we identified one genome-wide significant locus (rs144614292, chr11 : 20885143 G>T) located in the intron region of the gene *NELL1*, which codes the Neural EGFL-like protein 1 (NELL1). The colocalization analysis suggests that this region might be related to the blood mQTL (cg07126637) signal. Moreover, two more genes (*PROL1/OPRPN* and *MUC7*) from chr4 were identified to be gene-level *p*-value significant. The results of cell-type enrichment and functional analyses indicate that microglia are the most significantly enriched brain cell type, while immune response is the primary biological process associated with these genetic factors.

Our deep learning model accounts for cognitive decline contributed by normal aging

Our dual-loss Siamese ResNet model is grounded in a series of fundamental assumptions. 1) Instead of learning the supervised prediction problems such as AD versus CN, we assumed our deep learning model could learn the normal aging features from longitudinal CN MRI data by considering the outcome as a continuous metric to enhance its predictive power; 2) We hypothesized that such normal aging features would be distinguishable from AD-related MRI features (the magnitude of alteration in brain regions), therefore allowing us to disentangle AD-related cognitive decline from normal-aging-related cognitive decline; and 3) Within this model, we employed a dual loss framework incorporating both MSE and contrastive loss on pairs of longitudinal MRI and cognitive assessment scores. Conceptually, we postulated that the MSE loss applied to the initial time point of the pair would serve as a baseline, while the contrastive loss would ascertain whether the disparity in MRI images and the change in cognitive assessment scores exceed a predefined threshold for the normal aging effect.

To disentangle normal aging-related cognitive decline features from AD-related cognitive decline, we explicitly trained our model on a CN population and achieved –0.180 ± 0.261 (mean ± s.d.) on normalized predicted error in the validation set that comprised 946 MRI pairs from 125 subjects.

Our model demonstrated superior evaluation performance, with a more constrained dispersion of errors, when compared to two other designs of deep-learning models (ranking CNN and RNN), indicating that our model can effectively capture the subtle changes of MRI features related to normal aging by using longitudinal neuroimaging data. By employing the insights acquired from the trained model, we were able to differentiate between accelerated cognitive decline associated with AD and age-related cognitive decline, thus providing a more accurate depiction of accelerated cognitive decline with the capacity to better inform the genetic basis of accelerated cognitive decline in subsequent GWAS analyses.

Lastly, we observed a small increase in the slope of accelerated cognitive decline, occurring alongside the rise in the slope of normal aging-related cognitive decline within the CN and MCI groups. On the contrary, a much larger magnitude of increase was observed within the AD group (Fig. 3E). This trend enables the AD group (red) to diverge from the CN (green) and MCI (blue) groups, although all three groups exhibit similar accelerated cognitive decline within the low-end of normal-aging-related cognitive decline. Overall, we quantitatively depict the association between disentangled normal aging-related progress and disease-related progress among diagnosis groups.

Novel locus and genes

To understand the mechanism underlying this genome-wide significant locus, we explored if the same variant is responsible for the regulatory changes of genes among disease-relevant tissue-cell types. We adapted Bayesian-based Coloc analysis to identify the aligned evidence from publicly available quantitative trait locus (QTL) resources for the human brain and immune cell types. For the lead SNP (rs144614292) in chr11 locus, we identified posterior possibility H4 (0.086) for *NELL1* in excitatory neurons eQTL. For the lead SNP (rs4694308) in chr4 locus, cg07126637 (intron of SMR3B), we identified PP H4 (0.68), suggesting that such a "causal" relationship exists in human blood mQTL [81]; another CpG site, cg03970609 (intron of MUC7), does not show a significant colocalization signal. These findings warrant further aging-context brain evidence and experimental validation.

Our GWAS identified three genes related to accelerated cognitive decline, neither of which have been identified as related to AD or cognitive decline in previous GWAS, *NELL1*, *PROL1*, and *MCU7*; *SMR3B* was identified by the colocalization analysis. The novel locus in the intron of *NELL1* (rs144614292, $p = 3.73 \times 10^{-8}$) has been identified in our GWAS. *NELL1* encodes the protein NEL-like protein 1 (NELL1), a cytoplasmic protein that contains neural epidermal growth factor (EGF)-like repeats. *NELL1* has cytoplasmic expression in the brain, with low brain regional specificity, and is expressed mostly in oligodendrocytes precursor cells and in excitatory and inhibitory neurons. Accordingly, *NELL1* is involved in the modulation of synaptic plasticity via the regulation of its receptor CNTNAP4 (Contactin Associated Protein Family Member 4), which is crucial in synapse development [83]. *NELL1* has been found differentially expressed in the superior temporal gyrus (STG) and inferior frontal gyrus (IFG) of individuals with AD; the STG is a region showing atrophy and epigenetic changes, specifically in AD, while the IFG is a region in which atrophy is predominantly related to aging [85]. Interestingly, plasma levels of the protein encoded by *NELL1* are dysregulated in the earliest stage of AD, suggesting the protein coded by *NELL1* is a potential biomarker for early MCI and AD diagnosis [86]. The other two genes identified in our GWAS are *PROL1* and *MUC7*. Gene *PROL1*, also called *OPRPN*, encodes the protein opiorphin prepropeptide, a potent endogenous inhibitor of neprilysin, which crosses the blood-brain barrier [87]. Neprilysin is the central Aβ peptide-degrading enzyme in the brain, and it becomes down-regulated and inactive not only during the early stages of AD but also in normal aging. Thus, *PROL1* overexpression might be related to cognitive decline in general by inhibiting neprilysin and thus propitiating amyloid beta accumulation [88]. It has also been hypothesized that opiorphin might act as an antidepressant by activating both μ and δ opioid receptors indirectly [89]. Gene *MUC7* encodes the protein mucin-7 and has been implicated in cholesterol metabolism [90]. Increased serum levels of cholesterol have been identified as a risk factor for AD [91]. Gene *SMR3B* encodes the Submaxillary Gland Androgen Regulated Protein 3B, which is overexpressed in the salivary gland, testis, and pituitary from GTEx Portal [72]. Although *SMR3B* was identified as a significant PP in Coloc analysis for mQTL blood data, no direct evidence has yet linked *SMR3B* to cognitive decline or AD.

Our work has several limitations. First, our analyses are based on the assumption that normal aging MRI features differ from AD-related MRI features,

which might not necessarily be true. Previous studies [86, 87] have showed that AD-related and normal-aging-related cognitive decline would have shared regions of atrophy but different patterns and magnitudes, which would slightly reduce the accuracy of the predicted normal-aging-related cognitive decline. As shown in Fig. 3E, we indeed observe a positive association between the accelerated cognitive decline slope and predicted normal-aging-related cognitive decline slope within the AD group. In the future, we will explicitly use the non-overlapping regions of atrophy for training our model, as proposed by one recent study [92]. From the genetic correlation, we observed a weak positive correlation ($r_g = 0.10$, $p = 0.5$) between accelerated cognitive decline and education attainment, albeit not significant, which is not as we expected; it might have been raised from opposite effect directions across shared genetic variants, which might mask overall genetic correlation. Another limitation is that the age of the included participants was smaller for those who were CN than for those who had MCI or AD. The rs144614292-chr11 is a multiallelic SNP (G > A/G > T). We adapted its main genotype (G/A) in the GWAS analysis. However, this SNP is not recorded in most QTL databases, including GTEx, due to its multiallelic nature. Therefore, we only identified a weak H4 PP in excitatory neurons in a single-cell eQTL study [61]. We expect more solid associations will be identified with more comprehensive eQTL coverage. Lastly, due to the uniqueness of the ADNI dataset, no existing dataset has the same modalities. In the future, we will incorporate more datasets, such as ANMerge [93], and use the Z-score transformed-based method to make the clinical measurement comparable.

Conclusion remarks

Our new model has successfully extracted detailed information from MRI scans and was superior to cognitive evaluations alone. We quantitatively depicted the relationship between disentangled normal aging progression and disease-related advancement in diagnosis groups. We discovered a significant novel locus (rs144614292) situated in the intronic region of *NELL1*. A colocalization analysis pinpointed *SMR3B*, located in another locus with significant mapped genes *PROL1* and *MUC7*. Our technique exhibits promise in distinguishing accelerated cognitive decline from normal aging, pinpointing its genetic determinants, and providing improved prognostication and management of cognitive decline in patients. This paves the way for potential early intervention strategies.

ACKNOWLEDGMENTS

We would like to thank all the members of the Bioinformatics and Systems Medicine Laboratory (BSML) for constructive discussions.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research provides funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (https://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

FUNDING

We thank the support of the National Institutes of Health (NIH) grant U01AG079847 awarded to ZZ and XJ. We thanked the technical support from the Cancer Prevention and Research Institute of Texas (CPRIT RP180734). ZZ was partially supported

by NIH grants (R01LM012806, R01LM012806-07S1, and R03AG077191). AL was supported by a training fellowship from the Gulf Coast Consortia on Training in Precision Environmental Health Sciences (TPEHS) Training Grant (T32ES027801). AMM was supported by a training fellowship from the Gulf Coast Consortia on the NIH NLM Training Program in Biomedical Informatics & Data Science (T15LM007093). XJ is CPRIT Scholar in Cancer Research (RR180012), and he was partially supported by Christopher Sarofim Family Professorship, UT Stars award, UTHealth startup, NIH grants (R01AG066749, R01AG066749-03S1, U01AG079847, R01LM013712, and U01CA274576), and National Science Foundation (NSF #2124789).

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

DATA AVAILABILITY

The ADNI database is public for researchers and can be downloaded upon request at https://adni.loni.usc.edu (also see the Acknowledgments). All the data generated or analyzed in this study is available from the authors upon reasonable request. The overall framework can be downloaded from Github: https://github.com/davidroad/JAD.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JAD-231020.

REFERENCES

[1] (2021) 2021 Alzheimer's disease facts and figures. *Alzheimers Dement* **17**, 327-406.

[2] Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL (2006) Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry* **63**, 168-174.

[3] Sims R, Hill M, Williams J (2020) The multiplex model of the genetics of Alzheimer's disease. *Nat Neurosci* **23**, 311-322.

[4] Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, Boland A, Vronskaya M, van der Lee SJ, Amlie-Wolf A, et al. (2019) Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430.

[5] Pedersen NL, Gatz M, Berg S, Johansson B (2004) How heritable is Alzheimer's disease late in life? Findings from Swedish twins. *Ann Neurol* **55**, 180-185.

[6] Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, Rongve A, Børte S, Winsvold BS, Drange OK, Martinsen AE, Skogholt AH, Willer C, Bråthen G, Bosnes I, Nielsen JB, Fritsche LG, Thomas LF, Pedersen LM, Gabrielsen ME, Johnsen MB, Meisingset TW, Zhou W, Proitsi P, Hodges A, Dobson R, Velayudhan L, Heilbron K, Auton A, 23 and Me Research Team, Sealock JM, Davis LK, Pedersen NL, Reynolds CA, Karlsson IK, Magnusson S, Stefansson H, Thordardottir S, Jonsson PV, Snaedal J, Zettergren A, Skoog I, Kern S, Waern M, Zetterberg H, Blennow K, Stordal E, Hveem K, Zwart J-A, Athanasiu L, Selnes P, Saltvedt I, Sando SB, Ulstein I, Djurovic S, Fladby T, Aarsland D, Selbæk G, Ripke S, Stefansson K, Andreassen OA, Posthuma D (2021) A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* **53**, 1276-1282.

[7] Schwartzentruber J, Cooper S, Liu JZ, Barrio-Hernandez I, Bello E, Kumasaka N, Young AMH, Franklin RJM, Johnson T, Estrada K, Gaffney DJ, Beltrao P, Bassett A (2021) Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat Genet* **53**, 392-402.

[8] Hadjichrysanthou C, Evans S, Bajaj S, Siakallis LC, McRae-McKee K, de Wolf F, Anderson RM, Alzheimer's Disease Neuroimaging Initiative (2020) The dynamics of biomarkers across the clinical spectrum of Alzheimer's disease. *Alzheimers Res Ther* **12**, 74.

[9] Logue MW, Panizzon MS, Elman JA, Gillespie NA, Hatton SN, Gustavson DE, Andreassen OA, Dale AM, Franz CE, Lyons MJ, Neale MC, Reynolds CA, Tu X, Kremen WS (2019) Use of an Alzheimer's disease polygenic risk score to identify mild cognitive impairment in adults in their 50s. *Mol Psychiatry* **24**, 421-430.

[10] Arevalo-Rodriguez I, Smailagic N, Roqué-Figuls M, Ciapponi A, Sanchez-Perez E, Giannakou A, Pedraza OL, Bonfill Cosp X, Cullum S (2021) Mini-Mental State Examination (MMSE) for the early detection of dementia in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev* **7**, CD010783.

[11] Kueper JK, Speechley M, Montero-Odasso M (2018) The Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): Modifications and responsiveness in pre-dementia populations. A narrative review. *J Alzheimers Dis* **63**, 423-444.

[12] Devenney E, Hodges JR (2017) The Mini-Mental State Examination: Pitfalls and limitations. *Pract Neurol* **17**, 79-80.

[13] Ries ML, Carlsson CM, Rowley HA, Sager MA, Gleason CE, Asthana S, Johnson SC (2008) Magnetic resonance imaging characterization of brain structure and function in mild cognitive impairment: A review. *J Am Geriatr Soc* **56**, 920-934.

[14] Ledig C, Schuh A, Guerrero R, Heckemann RA, Rueckert D (2018) Structural brain imaging in Alzheimer's disease and mild cognitive impairment: Biomarker analysis and shared morphometry database. *Sci Rep* **8**, 11258.

[15] Stephen R, Liu Y, Ngandu T, Antikainen R, Hulkkonen J, Koikkalainen J, Kemppainen N, Lötjönen J, Levälahti E, Parkkola R, Pippola P, Rinne J, Strandberg T, Tuomilehto J, Vanninen R, Kivipelto M, Soininen H, Solomon A, FINGER study group (2019) Brain volumes and cortical thickness on MRI in the Finnish Geriatric Intervention Study

to Prevent Cognitive Impairment and Disability (FINGER). *Alzheimers Res Ther* **11**, 53.

[16] Counts SE, Ikonomovic MD, Mercado N, Vega IE, Mufson EJ (2017) Biomarkers for the early detection and progression of Alzheimer's disease. *Neurotherapeutics* **14**, 35-53.

[17] Hutton JS, Dudley J, Horowitz-Kraus T, DeWitt T, Holland SK (2020) Associations between screen-based media use and brain white matter integrity in preschool-aged children. *JAMA Pediatr* **174**, e193869.

[18] van Maurik IS, Vos SJ, Bos I, Bouwman FH, Teunissen CE, Scheltens P, Barkhof F, Frolich L, Kornhuber J, Wiltfang J, Maier W, Peters O, Rüther E, Nobili F, Frisoni GB, Spiru L, Freund-Levi Y, Wallin AK, Hampel H, Soininen H, Tsolaki M, Verhey F, Kłoszewska I, Mecocci P, Vellas B, Lovestone S, Galluzzi S, Herukka S-K, Santana I, Baldeiras I, de Mendonça A, Silva D, Chetelat G, Egret S, Palmqvist S, Hansson O, Visser PJ, Berkhof J, van der Flier WM, Alzheimer's Disease Neuroimaging Initiative (2019) Biomarker-based prognosis for people with mild cognitive impairment (ABIDE): A modelling study. *Lancet Neurol* **18**, 1034-1044.

[19] Fjell AM, McEvoy L, Holland D, Dale AM, Walhovd KB, Alzheimer's Disease Neuroimaging Initiative (2014) What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. *Prog Neurobiol* **117**, 20-40.

[20] Gonneaud J, Baria AT, Pichet Binette A, Gordon BA, Chhatwal JP, Cruchaga C, Jucker M, Levin J, Salloway S, Farlow M, Gauthier S, Benzinger TLS, Morris JC, Bateman RJ, Breitner JCS, Poirier J, Vachon-Presseau E, Villeneuve S, Alzheimer's Disease Neuroimaging Initiative (ADNI), Dominantly Inherited Alzheimer Network (DIAN) Study Group, Pre-symptomatic Evaluation of Experimental or Novel Treatments for Alzheimer's Disease (PREVENT-AD) Research Group (2021) Accelerated functional brain aging in pre-clinical familial Alzheimer's disease. *Nat Commun* **12**, 5346.

[21] Ouyang J, Zhao Q, Adeli E, Zaharchuk G, Pohl KM (2022) Disentangling normal aging from severity of disease via weak supervision on longitudinal MRI. *IEEE Trans Med Imaging* **41**, 2558-2569.

[22] Lorenzi M, Pennec X, Frisoni GB, Ayache N, Alzheimer's Disease Neuroimaging Initiative (2015) Disentangling normal aging from Alzheimer's disease in structural magnetic resonance images. *Neurobiol Aging* **36 Suppl 1**, S42-52.

[23] Veitch DP, Weiner MW, Aisen PS, Beckett LA, DeCarli C, Green RC, Harvey D, Jack CR Jr, Jagust W, Landau SM, Morris JC, Okonkwo O, Perrin RJ, Petersen RC, Rivera-Mindt M, Saykin AJ, Shaw LM, Toga AW, Tosun D, Trojanowski JQ, Alzheimer's Disease Neuroimaging Initiative (2022) Using the Alzheimer's Disease Neuroimaging Initiative to improve early detection, diagnosis, and treatment of Alzheimer's disease. *Alzheimers Dement* **18**, 824-857.

[24] van Loenhoud AC, van der Flier WM, Wink AM, Dicks E, Groot C, Twisk J, Barkhof F, Scheltens P, Ossenkoppele R, Alzheimer's Disease Neuroimaging Initiative (2019) Cognitive reserve and clinical progression in Alzheimer disease: A paradoxical relationship. *Neurology* **93**, e334-e346.

[25] Liu M, Zhang J, Adeli E, Shen D (2019) Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis. *IEEE Trans Biomed Eng* **66**, 1195-1206.

[26] Sarvamangala DR, Kulkarni RV (2022) Convolutional neural networks in medical image understanding: A survey. *Evol Intell* **15**, 1-22.

[27] Farooq A, Anwar S, Awais M, Rehman S (2017) A deep CNN based multi-class classification of Alzheimer's disease using MRI. In *2017 IEEE International Conference on Imaging Systems and Techniques (IST)* IEEE, pp. 1-6.

[28] AbdulAzeem Y, Bahgat WM, Badawy M (2021) A CNN based framework for classification of Alzheimer's disease. *Neural Comput Appl* **33**, 10415-10428.

[29] AlSaeed D, Omar SF (2022) Brain MRI analysis for Alzheimer's disease diagnosis using CNN-based feature extraction and machine learning. *Sensors (Basel)* **22**, 2911.

[30] Dyrba M, Hanzig M, Altenstein S, Bader S, Ballarini T, Brosseron F, Buerger K, Cantré D, Dechent P, Dobisch L, Düzel E, Ewers M, Fliessbach K, Glanz W, Haynes J-D, Heneka MT, Janowitz D, Keles DB, Kilimann I, Laske C, Maier F, Metzger CD, Munk MH, Perneczky R, Peters O, Preis L, Priller J, Rauchmann B, Roy N, Scheffler K, Schneider A, Schott BH, Spottke A, Spruth EJ, Weber M-A, Ertl-Wagner B, Wagner M, Wiltfang J, Jessen F, Teipel SJ, for the ADNI, AIBL, DELCODE study groups (2021) Improving 3D convolutional neural network comprehensibility via interactive visualization of relevance maps: Evaluation in Alzheimer's disease. *Alzheimers Res Ther* **13**, 191.

[31] Lei B, Liang E, Yang M, Yang P, Zhou F, Tan E-L, Lei Y, Liu C-M, Wang T, Xiao X, Wang S (2022) Predicting clinical scores for Alzheimer's disease based on joint and deep learning. *Expert Syst Appl* **187**, 115966.

[32] Qiao H, Chen L, Zhu F (2022) Ranking convolutional neural network for Alzheimer's disease mini-mental state examination prediction at multiple time-points. *Comput Methods Programs Biomed* **213**, 106503.

[33] 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. (2022) New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* **54**, 412-436.

[34] Mohs RC, Knopman D, Petersen RC, Ferris SH, Ernesto C, Grundman M, Sano M, Bieliauskas L, Geldmacher D, Clark C, Thal LJ (1997) Development of cognitive instruments for use in clinical trials of antidementia drugs: Additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord* **11 Suppl 2**, S13-S21.

[35] Routier A, Burgos N, Díaz M, Bacci M, Bottani S, El-Rifai O, Fontanella S, Gori P, Guillon J, Guyot A, Hassanaly R, Jacquemont T, Lu P, Marcoux A, Moreau T, Samper-González J, Teichmann M, Thibeau-Sutre E, Vaillant G, Wen J, Wild A, Habert M-O, Durrleman S, Colliot O (2021) Clinica: An open-source software platform for reproducible clinical neuroscience studies. *Front Neuroinform* **15**, 689675.

[36] Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, Flandin G, Ghosh SS, Glatard T, Halchenko YO, Handwerker DA, Hanke M, Keator D, Li X, Michael Z, Maumet C, Nichols BN, Nichols TE, Pellman J, Poline J-B, Rokem A, Schaefer G, Sochat V, Triplett W, Turner JA, Varoquaux G, Poldrack RA (2016) The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data* **3**, 160044.

[37] Sled JG, Zijdenbos AP, Evans AC (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* **17**, 87-97.

[38] Valverde S, Coll L, Valencia L, Clèrigues A, Oliver A, Vilanova JC, Ramió-Torrentà L, Rovira À, Lladó X (2021) Assessing the accuracy and reproducibility of PARIETAL: A deep learning brain extraction algorithm. *J Magn Reson Imaging*, doi: 10.1002/jmri.27776.

[39] Fischl B (2012) FreeSurfer. *Neuroimage* **62**, 774-781.

[40] Schaeverbeke JM, Gabel S, Meersmans K, Luckett ES, De Meyer S, Adamczuk K, Nelissen N, Goovaerts V, Radwan A, Sunaert S, Dupont P, Van Laere K, Vandenberghe R (2021) Baseline cognition is the best predictor of 4-year cognitive change in cognitively intact older adults. *Alzheimers Res Ther* **13**, 75.

[41] McDonough IM, Popp TE (2020) Linear and nonlinear relationships between cognitive subdomains of ability discrepancy and Alzheimer's disease biomarkers. *Neuropsychology* **34**, 211-226.

[42] Mattsson-Carlgren N, Salvadó G, Ashton NJ, Tideman P, Stomrud E, Zetterberg H, Ossenkoppele R, Betthauser TJ, Cody KA, Jonaitis EM, Langhough R, Palmqvist S, Blennow K, Janelidze S, Johnson SC, Hansson O (2023) Prediction of longitudinal cognitive decline in preclinical Alzheimer disease using plasma biomarkers. *JAMA Neurol* **80**, 360-369.

[43] Sherva R, Tripodis Y, Bennett DA, Chibnik LB, Crane PK, de Jager PL, Farrer LA, Saykin AJ, Shulman JM, Naj A, Green RC, GENAROAD Consortium, Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Genetics Consortium (2014) Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimers Dement* **10**, 45-52.

[44] Chicco D (2021) Siamese neural networks: An overview. *Methods Mol Biol* **2190**, 73-94.

[45] Bromley J, Guyon I, LeCun Y, Säckinger E, Shah R (1993) Signature verification using a "Siamese" time delay neural network. *Adv Neural Inf Process Syst* **6**, 737-744.

[46] He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778.

[47] Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. *arXiv [csLG]*.

[48] Alzheimer's Disease Neuroimaging Initiative Database, ADNI database.

[49] Li X, Fernandes BS, Liu A, Lu Y, Chen J, Zhao Z, Dai Y (2023) Genetically-regulated pathway-polygenic risk score (GRPa-PRS): A risk stratification method to identify genetically regulated pathways in polygenic diseases. *medRxiv*, https://doi.org/10.1101/2023.06.19.23291621 [Preprint]. Posted June 27, 2023.

[50] Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ (2006) The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Res* **34**, D590-8.

[51] McCarthy Group Tools, https://www.chg.ox.ac.uk/~wrayner/tools/

[52] Michigan Imputation Server, https://imputationserver.sph.umich.edu/index.html#!.

[53] Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873.

[54] Wang K, Li M, Hakonarson H (2010) ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164.

[55] Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993.

[56] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.

[57] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575.

[58] IGSR: The International Genome Sample Resource, 1000 Genome Project ftp.

[59] McCaw ZR, Lane JM, Saxena R, Redline S, Lin X (2020) Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics* **76**, 1262-1272.

[60] Zheng Z, Huang D, Wang J, Zhao K, Zhou Y, Guo Z, Zhai S, Xu H, Cui H, Yao H, Wang Z, Yi X, Zhang S, Sham PC, Li MJ (2020) QTLbase: An integrative resource for quantitative trait loci across multiple human molecular phenotypes. *Nucleic Acids Res* **48**, D983-D991.

[61] Bryois J, Calini D, Macnair W, Foo L, Urich E, Ortmann W, Iglesias VA, Selvaraj S, Nutma E, Marzin M, Amor S, Williams A, Castelo-Branco G, Menon V, De Jager P, Malhotra D (2022) Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat Neurosci* **25**, 1104-1112.

[62] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383.

[63] Wang G, Sarkar A, Carbonetto P, Stephens M (2020) A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Series B Stat Methodol* **82**, 1273-1300.

[64] Manuel AM, Dai Y, Jia P, Freeman LA, Zhao Z (2023) A gene regulatory network approach harmonizes genetic and epigenetic signals and reveals repurposable drug candidates for multiple sclerosis. *Hum Mol Genet* **32**, 998-1009.

[65] Manuel AM, Dai Y, Freeman LA, Jia P, Zhao Z (2021) An integrative study of genetic variants with brain tissue expression identifies viral etiology and potential drug targets of multiple sclerosis. *Mol Cell Neurosci* **115**, 103656.

[66] Gagliano Taliun SA, VandeHaar P, Boughton AP, Welch RP, Taliun D, Schmidt EM, Zhou W, Nielsen JB, Willer CJ, Lee S, Fritsche LG, Boehnke M, Abecasis GR (2020) Exploring and visualizing large-scale genetic associations by using PheWeb. *Nat Genet* **52**, 550-552.

[67] de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015) MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219.

[68] Watanabe K, Taskesen E, van Bochoven A, Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826.

[69] Watanabe K, Umićević Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D (2019) Genetic mapping of cell type specificity for complex traits. *Nat Commun* **10**, 3222.

[70] Fernandes B, Enduru N, Fernandes B, Bahrami S, Dai Y, Andreassen O, Zhao Z (2023) Genetic overlap between Alzheimer's disease and immune-mediated diseases: An atlas of shared genetic determinants and biological convergence. *Research Square*, doi: 10.21203/rs.3.rs-3346282/v1 [Preprint]. Posted September 28, 2023.

[71] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425.

[72] GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585.

[73] Dai Y, Hu R, Liu A, Cho KS, Manuel AM, Li X, Dong X, Jia P, Zhao Z (2022) WebCSEA: Web-based cell-type-specific enrichment analysis of genes. *Nucleic Acids Res* **50**, W782-W790.

[74] Pei G, Dai Y, Zhao Z, Jia P (2019) deTS: Tissue-specific enrichment analysis to decode tissue specificity. *Bioinformatics* **35**, 3842-3845.

[75] Dai Y, Hu R, Manuel AM, Liu A, Jia P, Zhao Z (2021) CSEA-DB: An omnibus for human complex trait and cell type associations. *Nucleic Acids Res* **49**, D862-D870.

[76] Pei G, Hu R, Dai Y, Manuel AM, Zhao Z, Jia P (2021) Predicting regulatory variants using a dense epigenomic mapped CNN model elucidated the molecular basis of trait-tissue associations. *Nucleic Acids Res* **49**, 53-66.

[77] Pei G, Hu R, Jia P, Zhao Z (2021) DeepFun: A deep learning sequence-based model to decipher non-coding variant effect in a tissue- and cell type-specific manner. *Nucleic Acids Res* **49**, W131-W139.

[78] Privé F, Arbel J, Vilhjálmsson BJ (2021) LDpred2: Better, faster, stronger. *Bioinformatics* **36**, 5424-5431.

[79] Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC (2018) The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408.

[80] Liu A, Manuel AM, Dai Y, Zhao Z (2022) Prioritization of risk genes in multiple sclerosis by a refined Bayesian framework followed by tissue-specificity and cell type feature assessment. *BMC Genomics* **23**, 362.

[81] Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, Carnero-Montoro E, Lawson DJ, Burrows K, Suderman M, Bretherick AD, Richardson TG, Klughammer J, Iotchkova V, Sharp G, Al Khleifat A, Shatunov A, Iacoangeli A, McArdle WL, Ho KM, Kumar A, Söderhäll C, Soriano-Tárraga C, Giralt-Steinhauer E, Kazmi N, Mason D, McRae AF, Corcoran DL, Sugden K, Kasela S, Cardona A, Day FR, Cugliari G, Viberti C, Guarrera S, Lerro M, Gupta R, Bollepalli S, Mandaviya P, Zeng Y, Clarke T-K, Walker RM, Schmoll V, Czamara D, Ruiz-Arenas C, Rezwan FI, Marioni RE, Lin T, Awaloff Y, Germain M, Aïssi D, Zwamborn R, van Eijk K, Dekker A, van Dongen J, Hottenga J-J, Willemsen G, Xu C-J, Barturen G, Català-Moll F, Kerick M, Wang C, Melton P, Elliott HR, Shin J, Bernard M, Yet I, Smart M, Gorrie-Stone T, BIOS Consortium, Shaw C, Al Chalabi A, Ring SM, Pershagen G, Melén E, Jiménez-Conde J, Roquer J, Lawlor DA, Wright J, Martin NG, Montgomery GW, Moffitt TE, Poulton R, Esko T, Milani L, Metspalu A, Perry JRB, Ong KK, Wareham NJ, Matullo G, Sacerdote C, Panico S, Caspi A, Arseneault L, Gagnon F, Ollikainen M, Kaprio J, Felix JF, Rivadeneira F, Tiemeier H, van IJzendoorn MH, Uitterlinden AG, Jaddoe VWV, Haley C, McIntosh AM, Evans KL, Murray A, Räikkönen K, Lahti J, Nohr EA, Sørensen TIA, Hansen T, Morgen CS, Binder EB, Lucae S, Gonzalez JR, Bustamante M, Sunyer J, Holloway JW, Karmaus W, Zhang H, Deary IJ, Wray NR, Starr JM, Beekman M, van Heemst D, Slagboom PE, Morange P-E, Trégouët D-A, Veldink JH, Davies GE, de Geus EJC, Boomsma DI, Vonk JM, Brunekreef B, Koppelman GH, Alarcón-Riquelme ME, Huang R-C, Pennell CE, van Meurs J, Ikram MA, Hughes AD, Tillin T, Chaturvedi N, Pausova Z, Paus T, Spector TD, Kumari M, Schalkwyk LC, Visscher PM, Davey Smith G, Bock C, Gaunt TR, Bell JT, Heijmans BT, Mill J, Relton CL (2021) Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet* **53**, 1311-1321.

[82] Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N, Daly MJ, Price AL, Neale BM (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295.

[83] Li C, Zheng Z, Ha P, Chen X, Jiang W, Sun S, Chen F, Asatrian G, Berthiaume EA, Kim JK, Chen EC, Pang S, Zhang X, Ting K, Soo C (2018) Neurexin superfamily cell membrane receptor contactin-associated protein like-4 (Cntnap4) is involved in neural EGFL-like 1 (Nell-1)-responsive osteogenesis. *J Bone Miner Res* **33**, 1813-1825.

[84] Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen G-B, Emilsson V, Meddens SFW, et al. (2016) Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-542.

[85] Li QS, De Muynck L (2021) Differentially expressed genes in Alzheimer's disease highlighting the roles of microglia genes including OLR1 and astrocyte gene CDK2AP1. *Brain Behav Immun Health* **13**, 100227.

[86] Jiang Y, Zhou X, Ip FC, Chan P, Chen Y, Lai NCH, Cheung K, Lo RMN, Tong EPS, Wong BWY, Chan ALT, Mok VCT, Kwok TCY, Mok KY, Hardy J, Zetterberg H, Fu AKY, Ip NY (2022) Large-scale plasma proteomic profiling identifies a high-performance biomarker panel for Alzheimer's disease screening and staging. *Alzheimers Dement* **18**, 88-102.

[87] Wisner A, Dufour E, Messaoudi M, Nejdi A, Marcel A, Ungeheuer M-N, Rougeot C (2006) Human Opiorphin, a natural antinociceptive modulator of opioid-dependent pathways. *Proc Natl Acad Sci U S A* **103**, 17979-17984.

[88] El-Amouri SS, Zhu H, Yu J, Marr R, Verma IM, Kindy MS (2008) Neprilysin: An enzyme candidate to slow the progression of Alzheimer's disease. *Am J Pathol* **172**, 1342-1354.

[89] Yang Q-Z, Lu S-S, Tian X-Z, Yang A-M, Ge W-W, Chen Q (2011) The antidepressant-like effect of human opiorphin via opioid-dependent pathways in mice. *Neurosci Lett* **489**, 131-135.

[90] Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, Surakka I, Ntalla I, Vedantam S, Winkler TW, et al. (2021) The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675-679.

[91] Loera-Valencia R, Goikolea J, Parrado-Fernandez C, Merino-Serrais P, Maioli S (2019) Alterations in cholesterol metabolism as a risk factor for developing Alzheimer's dis-

ease: Potential novel targets for treatment. *J Steroid Biochem Mol Biol* **190**, 104-114.

[92] Hwang G, Abdulkadir A, Erus G, Habes M, Pomponio R, Shou H, Doshi J, Mamourian E, Rashid T, Bilgel M, Fan Y, Sotiras A, Srinivasan D, Morris JC, Albert MS, Bryan NR, Resnick SM, Nasrallah IM, Davatzikos C, Wolk DA, from the iSTAGING consortium, ADNI (2022) Disentangling Alzheimer's disease neurodegeneration from typical brain ageing using machine learning. *Brain Commun* **4**, fcac117.

[93] Birkenbihl C, Westwood S, Shi L, Nevado-Holgado A, Westman E, Lovestone S, AddNeuroMed Consortium, Hofmann-Apitius M (2021) ANMerge: A comprehensive and accessible Alzheimer's disease patient-level dataset. *J Alzheimers Dis* **79**, 423-431.