Xu *et al. BioData Mining* (2025) 18:24
https://doi.org/10.1186/s13040-025-00432-1

BioData Mining

RESEARCH Open Access

High-dimensional mediation analysis reveals the mediating role of physical activity patterns in genetic pathways leading to AD-like brain atrophy

Hanxiang Xu[1,3], Shizhuo Mu[1], Jingxuan Bao[1], Christos Davatzikos[1,2], Haochang Shou[1*] and Li Shen[1,2*]

*Correspondence: haochang.shou@pennmedicine.upenn.edu; li.shen@pennmedicine.upenn.edu

[1] Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[2] Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[3] Department of Family Medicine and Public Health, University of California, San Diego, CA 92093, USA

Abstract

Background: Alzheimer's disease (AD) is a complex disorder that affects multiple biological systems including cognition, behavior and physical health. Unfortunately, the pathogenic mechanisms behind AD are not yet clear and the treatment options are still limited. Despite the increasing number of studies examining the pairwise relationships between genetic factors, physical activity (PA), and AD, few have successfully integrated all three domains of data, which may help reveal mechanisms and impact of these genomic and phenomic factors on AD. We use high-dimensional mediation analysis as an integrative framework to study the relationships among genetic factors, PA and AD-like brain atrophy quantified by spatial patterns of brain atrophy.

Results: We integrate data from genetics, PA and neuroimaging measures collected from 13,425 UK Biobank samples to unveil the complex relationship among genetic risk factors, behavior and brain signatures in the contexts of aging and AD. Specifically, we used a composite imaging marker, Spatial Pattern of Abnormality for Recognition of Early AD (SPARE-AD) that characterizes AD-like brain atrophy, as an outcome variable to represent AD risk. Through GWAS, we identified single nucleotide polymorphisms (SNPs) that are significantly associated with SPARE-AD as exposure variables. We employed conventional summary statistics and functional principal component analysis to extract patterns of PA as mediators. After constructing these variables, we utilized a high-dimensional mediation analysis method, Bayesian Mediation Analysis (BAMA), to estimate potential mediating pathways between SNPs, multivariate PA signatures and SPARE-AD. BAMA incorporates Bayesian continuous shrinkage prior to select the active mediators from a large pool of candidates. We identified a total of 22 mediation pathways, indicating how genetic variants can influence SPARE-AD by altering physical activity. By comparing the results with those obtained using univariate mediation analysis, we demonstrate the advantages of high-dimensional mediation analysis methods over univariate mediation analysis.

Conclusion: Through integrative analysis of multi-omics data, we identified several mediation pathways of physical activity between genetic factors and SPARE-AD. These findings contribute to a better understanding of the pathogenic mechanisms of AD.

Moreover, our research demonstrates the potential of the high-dimensional mediation analysis method in revealing the mechanisms of disease.

Keywords: High-dimensional mediation analysis, Physical activity, Genetic risk factors, Alzheimer's disease, SPARE-AD index, Disease mechanism

Background

As the population ages, increasing research efforts are devoted to studying human aging process and age-related diseases such as neurodegenerations [1–4]. Alzheimer's disease and related dementias (ADRD) remain formidable challenges in the public health landscape. The disease has demonstrated profound societal and economic impact, affecting millions globally and leading to significant public health burdens. As of 2023, an estimated 6.7 million Americans aged 65 and older are living with Alzheimer's dementia, underlining its status as one of the costliest conditions to the society [5]. This prevalence is projected to escalate, compounding the urgency for effective interventions. However, to this date, though there are numerous AD-related studies [6–8], the pathogenic mechanisms for AD remain not well understood, and little progress has been made for identifying effective solutions for treating and managing the disease. Given the complexity and heterogeneity of how the disease affects human body, it might be necessary to integrate multimodal and multi-omics measures when revealing the biological mechanisms and identifying potential targets for therapeutics [9].

Recent studies have shown physical activity (PA) patterns are heritable traits and are correlated with several known genetic risk factors of AD including APOE gene, the best-known gene associated with AD [10]. Some studies suggest that increased physical movement might be beneficial reducing AD risk [11–14]. However, due to the lack of appropriate data and statistical methods for handling complex multi-omics data, few studies have directly linked physical activity with well-known AD-related risk factors and biomarkers such as genetic variants brain changes, and extensively evaluate the relationship of the three.

Mediation analysis has emerged as one of the powerful and increasingly popular tools in biomedical studies and clinical research [15–21]. It enables unraveling of the mechanisms and pathways through which causal effects operate. In our setting, we have a group of genetic risk factors to be considered as exposure variables. Additionally, there might be a group of potentially high-dimensional mediators that could reside on the pathway between each exposure and outcome variable. This poses analytic challenges that could not be addressed by the classical univariate mediation analysis. High-dimensional mediation modeling techniques that account for correlations among multiple mediators and identify significant mediating effects are desirable [22–25]. In this paper, we utilize the recent advances of high-dimensional mediation analysis methodologies to investigate the joint relationship among genetics, physical activity and AD-related neuroimaging markers [26–32]. Figure 1 shows a potential mediation relationship: physical activity might mediate the effect of genetic variation on AD-like brain atrophy index. By identifying such an effect, we can better understand the mechanisms of action among these three factors, providing recommendations and insights for treating or mitigating the progression of AD

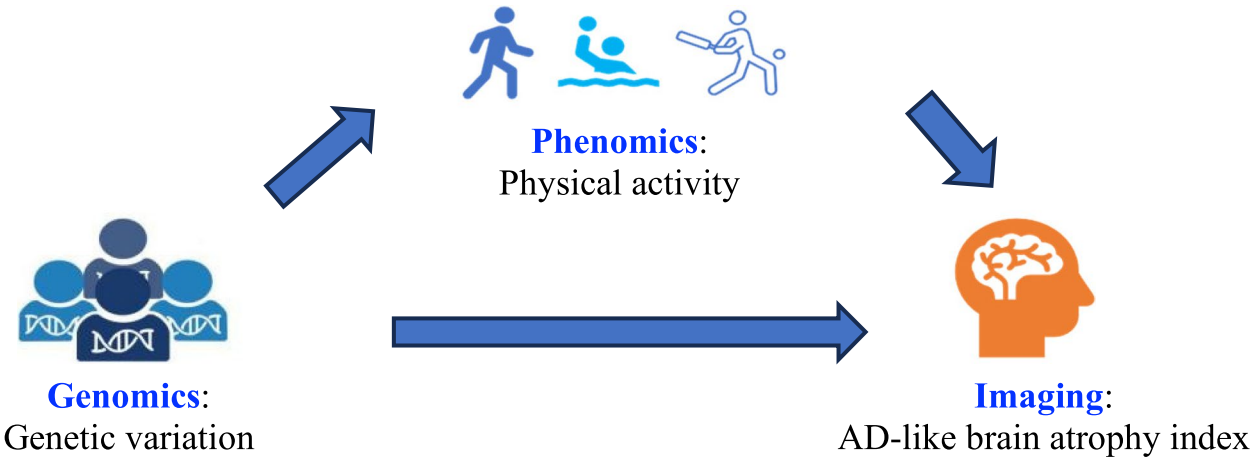

Fig. 1 Physical activity might mediate the effect of genetic variation on AD-like brain atrophy index

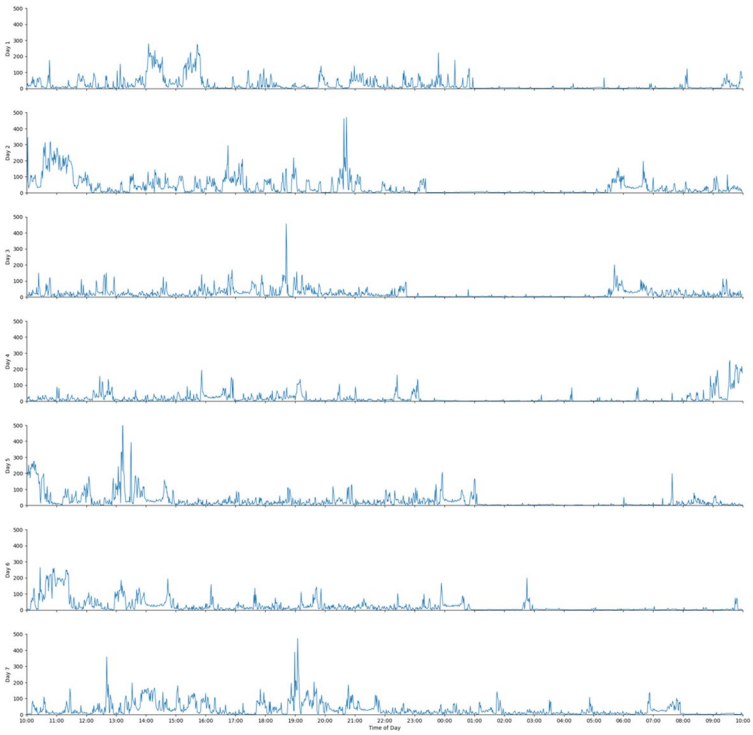

Fig. 2 Graph of activity counts over time of a certain participant. The Y-axis represents activity counts, and the X-axis represents time (from 10:00 AM on the first day to 10:00 AM on the second day)

Methods

Study population

The UK Biobank (UKBB) is a large prospective cohort study which enrolled more than 500,000 individuals aged from 37 to 70 years with approximately 88% having British ancestry [33]. The UK Biobank collected an exceptional breadth and depth of information on various factors including sociodemographic, lifestyle, environment, accelerometry, imaging, and genetics. Participants were recruited from the United Kingdom with initial enrollment carried out from 2006 to 2010. Our study included UKBB samples with genetic data, structural magnetic resonance imaging (MRI) data used for calculating the AD-like brain atrophy score, and physical activity data recorded from accelerometers.

Physical activity data

Physical activity measures human behavior and activity levels, which are related to the effects of genetic variants and Alzheimer's disease on individuals [10, 13]. In our study, we extracted features from physical activity data which were collected from a subset of UKBB participants using tri-axial wrist-worn accelerometers for up to 7 days [34] based on previous literature using multiple approaches as mediators. We included the physical activity data from 17,998 subjects who also have structural MRI collected. Figure 2 displays the daily epoch-level physical activity intensities of an example subject from 10:00am to 9:59am (next day) over a continuous seven-day period. The raw data consist of the average acceleration measured in 5-second intervals for each individual, which

provides sufficient resolution for distinguishing different types of physical activity (sedentary, light, moderate, vigorous).

We further aggregated the 5-second level data into minute-level resolution by calculating the average of 5-second interval within a minute to reduce the computation cost. We applied similar exclusion criterion as in Leroux et al. [35] to ensure data quality and apply the same pipeline for extracting a vector of physical activity features.

Physical activity features

We extracted two types of physical activity features, conventional summary statistics and principal component (PC) scores from functional principal component analysis (PCA). The set of conventional summary statistics are interpretable and commonly used, which include total acceleration (TA), total log (1 + acceleration) which is labeled total log acceleration (TLA), TLA in 2-hour windows, total sedentary time (ST), where sedentary is defined for each minute if the average milli-g in a particular minute below a given threshold, and others. There are 27 summary statistics in total; and Table 1 shows a few examples.

These features are derived for each day of every subject's observation. To aggregate across multiple days, we then calculated the mean and standard deviation (SD) of each summary statistics across days, resulting in a total of 54 features per subject.

Although conventional summary statistics are easy to obtain and understand, they might result in a loss of information due to the radical data compression. Hence, we further perform data-driven feature extraction using functional PCA (FPCA) to better capture complex dependency structures in the time series of high-resolution activity intensities. FPCA is an extension of traditional PCA method in the functional data analysis field where the data objects are continuous functions or curves of time or space, rather than finite-dimensional vectors. Such approaches have been widely adapted to analyze patterns of physical activity data [34, 35].

To mitigate the significant skewness in the activity count data, we first transform the minute-level activity intensity using the transformation $x = log(1 + a)$, where a represents the activity count. This transformation also ensures that zero counts remain zero. We then apply FPCA to obtain a set of PC scores for each subject each day. Let J_i represent number of days of accelerometry data for subject $i = 1, \ldots, N$ and let J be the total number of days of data. The log-transformed activity count data matrix, X, is $J \times 1440$ dimensional. We use the fast covariance estimation (FACE) approach [36] implemented in the 'fpca.face()' function of the 'refund' [37] package in R to PC estimations efficiently for high dimensional data. Subsequently, we projected each day's activity intensity data onto the first K PCs (see Fig. 3 for a few examples) and calculated the corresponding principal scores. More specifically, let c_{ijk} be the score for subject i, on day $j$and principal component k. We then construct 2 K variables by computing the mean and standard deviation of these subject-specific scores:

$$m_{ik} = \bar{c_{ik}} = \frac{1}{J}\sum_{j=1}^{J_i} c_{ikj}, s_{ik} = sd(c_{ik}) = \sqrt{\frac{\sum_{j=1}^{J_i}\left(c_{ijk} - \bar{c_{ik}}\right)^2}{J_i - 1}}, i = 1, \ldots, Nk = 1, \ldots, K = 50$$

Table 1 Conventional Summary statistics. Note that TLAC (1-12) represents the total log acceleration within twelve 2-hour windows: 10am-12pm, 12pm-2pm, 2pm-4pm, 4pm-6pm, 8pm-10pm, 10pm-12am, 12am-2am, 2am-4am, 4am-6am, 6am-8am, and 8am-10am

Summary statistics	Description
TAC	Total acceleration
TLAC	Total log acceleration
TLAC (1–12)	Total log acceleration during two-hour windows
ST	Sedentary/sleep minutes
LIPA	Light-intensity physical activity minutes
MVPA	Moderate-to-vigorous activity minutes
DARE	Daytime activity ratio estimate
SBout	Average duration of continuous sedentary or sleep periods for each day
ABout	Average duration (in minutes) of continuous active periods for each day
SATP	Sedentary/sleep to active transition probability
ASTP	Active to sedentary/sleep transition probability
Timing of M10	Mid-point of the ten most active hours of the day
Timing of L5	Mid-point of the five least active hours of the day
M10	Average log acceleration during the ten most active hours of the day
L5	Average log acceleration during the five least active hours of the day
RA	Relative amplitude

We derived 50 principal components from FPCA and got 100 features from this approach, including 50 from the mean measures and 50 from the standard deviation measures. Hence a total of 154 features are generated using conventional and data-driven feature extraction methods from accelerometry data, which then served as the potential mediator variables.

Imaging data

The imaging outcome used in our study is a composite brain atrophy biomarker, SPARE-AD (Spatial Pattern of Abnormality for Recognition of Early Alzheimer's disease) index [38], derived from volumetric measures of structural MRI data from UK Biobank study. The MRI scans were processed and harmonized using standards from the Imaging-Based Coordinate System for Aging and Neurodegenerative Diseases (iSTAGING) consortium [39]. The images underwent magnetic field intensity inhomogeneity correction [40] and brain segmentation using MUSE [41], a MUlti-atlas Segmentation method that utilizes Ensembles of registration algorithms and parameters along with locally optimal atlas selection. Volumes of region-of-interest are then calculated based on the MUSE segmentation as input for deriving SPARE-AD. More details about image preprocessing are available in [40].

SPARE-AD index

Machine learning-based aging indexes have emerged as powerful tools in aging research, providing a detailed understanding of the aging process beyond what traditional measures offer [42–45]. We adopted SPARE-AD (Spatial Pattern of Abnormality for Recognition of Early Alzheimer's disease) index [38] to quantify AD-like brain atrophy. The SPARE-AD index is a neuroimaging biomarker tool developed to identify early stage of AD by capturing spatial patterns of brain atrophy associated with the disease. It aids in

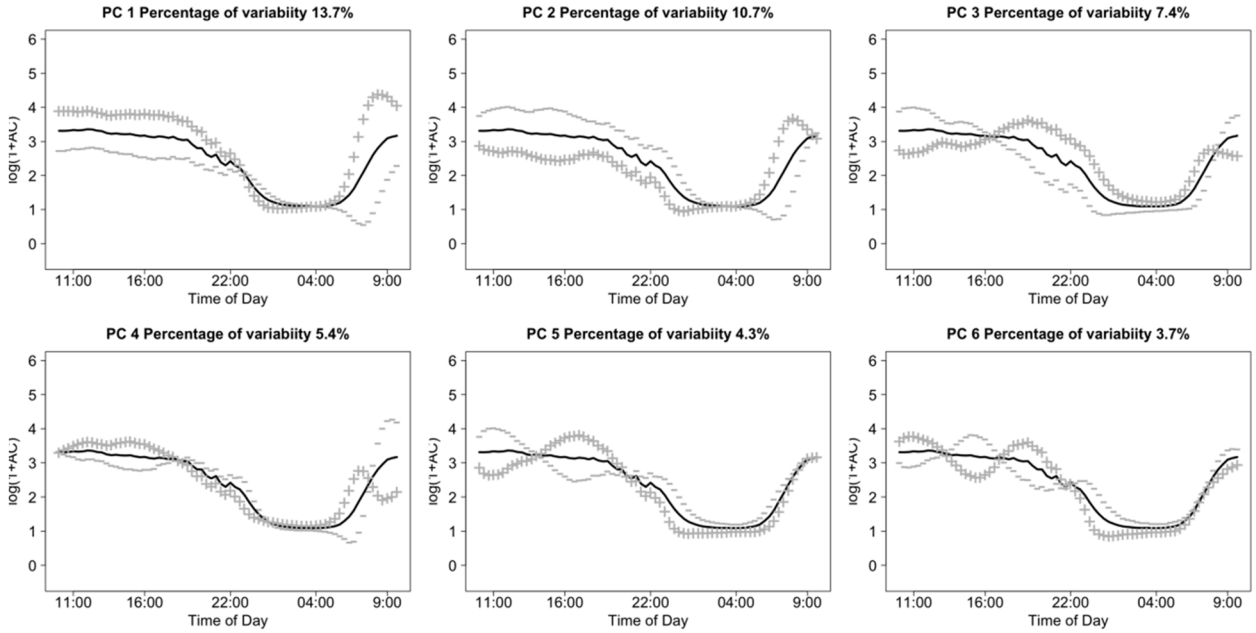

Fig. 3 Principal components (PCs) calculated on the population, minute level UKBB accelerometry data using functional principal component analysis (FPCA). Solid lines represent the population average curve, dashed lines with + and - signs denote the activity patterns when adding or subtracting a specific principal component with 2 standard deviations of PC score to the average activity curve. The title of each panels explains the percentage of variations explained by each PC. PCs 1 and 2 accounts for the most of data variability of 13.7% and 10.7%. PC1 demonstrates the variability in physical activity patterns among subjects where certain subjects have higher than average activity levels, especially during earlier morning hours and afternoon to evening period. PC2 shows subjects with higher than the average activity levels in the afternoon but might be a late riser than others

distinguishing between individuals with cognitively normal (CN), mild cognitive impairment (MCI), and AD, as well as predicting progression from CN to MCI and from MCI to AD by quantifying AD-related patterns of brain atrophy. This index has been widely used in AD-related studies and has shown great performance in predicting AD risk [46, 47]. In our sample, positive SPARE-AD index accounts for 4.2%, while negative SPARE-AD index accounts for 95.8%, as shown in Fig. 4. SPARE-AD index was computed using the imaging data from UKBB, and the machine learning model is based on previous independent studies [38, 40, 48].

Genetic data

Genetic data are sourced from the UKBB and processed according to established protocols [49]. The preprocessing data pipeline includes imputation and quality control (QC). Initially, subjects related to the second degree or closer were removed. The data were then refined by excluding multiallelic variants, variants with more than 3% missing call rates, those with minor allele frequencies below 1% and variants not meeting the Hardy-Weinberg equilibrium with a p-value threshold of 1e-10. Further filtering excluded subjects with missing call rates over 3% and those whose heterozygosity rate deviated five standard deviations from the norm. The final step involved synchronizing the quality controlled (QCed) imputed genotyping data with the QCed imaging data. The resulting UKBB imputed genetic dataset consisted of 482,831 SNPs and 38,195 subjects, which was subsequently used in our GWAS. Additionally, the first 50 genetic principal components (PCs) were derived for further analysis.

Genetic variants selection

To identify the genetic variants for our study, we used SAIGE [50] to filter out the genetic variants significantly associated with the SPARE-AD index. Scalable and Accurate Implementation of GEneralized mixed model (SAIGE) is a statistical tool designed for large-scale association studies of complex traits using mixed models and has shown great performance in many studies [51, 52]. It addresses challenges such as population stratification and relatedness among individuals which are common in traditional mixed model analyses for GWAS. We downloaded a total of 1,048,575 SNPs, setting the p-value threshold at 5e-8. The covariates included in our analyses were sex, age, body mass index (BMI) and 10 first genetic principal components. Table 2 summarized the sample characteristics and the distributions of their physical activity features of the 13,425 participants included in our mediation analysis, stratified by SPARE-AD positive and negative.

High-dimensional mediation analysis

We used BAyesian Mediation Analysis (BAMA), developed by Song et al. in 2020 [24, 53], to identify active mediators from a large pool of candidate mediators. BAMA incorporates a Bayesian continuous shrinkage prior to identify active mediators in the high-dimensional mediation analysis method (Fig. 5), This method could accommodate one exposure, one outcome and multiple mediators.

There are two primary models in this high-dimensional mediation analysis method: the outcome model and the mediator model.

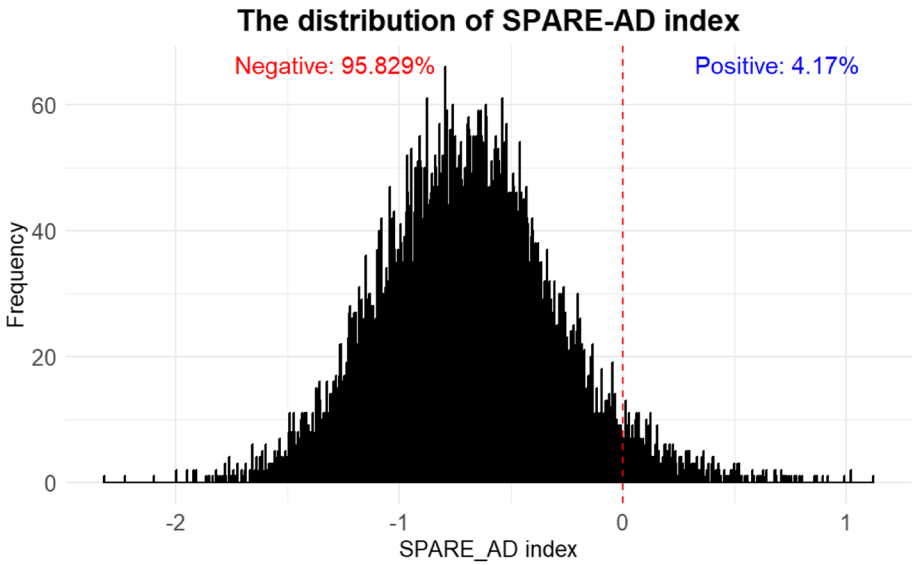

Fig. 4 Distribution of SPARE-AD indices, all the subjects are assumed to be cognitively normal. More positive SPARE-AD index indicates a higher AD risk, while more negative values imply lower AD risk

Table 2 Sample characteristics of the UKBB data used in our study. We used t-test to test the group difference for BMI, Chi-square test to test the group difference for sex

Characteristic	Positive $n_1 = 560$	Negative $n_2 = 12{,}865$	Total $n = 13{,}425$	*p*-value*
Age, median (range)	71.57(45.65–81.82)	64.52(45.46–81.19)	64.83(45.46–81.82)	1.50E-97
Sex: Female, No. (%)	187(33)	7172(56)	7359(55)	3.67E-25
Sex: Male, No. (%)	373(67)	5693(44)	6066(45)	
BMI, mean (SD)	26.26(4.13)	26.09(3.98)	26.10(3.98)	0.44
TAC_mean, mean (SD)	39,079(10596)	41,603(12144)	41,498(12094)	3.29E-06
TAC_sd, mean(SD)	7970(4500)	9042(6460)	8997(6393)	7.06E-05
MVPA_mean, mean(SD)	89(45)	99(51)	99(51)	2.07E-05
MVPA_sd, mean(SD)	37(19)	39(21)	39(21)	3.00E-02
PC1_mean, mean(SD)	0.2677(9.21)	0.0063(9.43)	0.0172(9.42)	0.52
PC1_sd, mean(SD)	8.13(3.12)	8.05(3.12)	8.06(3.12)	0.57
SPARE-AD index	0.21(0.20)	−0.73(0.34)	−0.69(0.39)	0.00E + 00

**p*-values are calculated by: Sex, Chi-square test; BMI, t-test; Age, TAC_mean, TAC_sd, MVPA_mean, MVPA_sd, PC1_mean, PC1_sd, Wilcoxon rank-sum test; SPARE-AD index, Kolmogorov-Smirnov test

1) Outcome model:

$$\mathrm{Y_i} = \mathbf{M_i^T \beta_m} + \mathrm{A_i}\beta_{\mathrm{a}} + \mathbf{C_i^T \beta_c} + \varepsilon_{\mathrm{Y_i}}$$

$$\beta_{\mathrm{m}} = \left((\beta_{\mathrm{m}})_1, \ldots, (\beta_{\mathrm{m}})_{\mathrm{p}}\right)^{\mathrm{T}}, \beta_{\mathrm{c}} = \left(\beta_{\mathrm{c1}}, \ldots, \beta_{\mathrm{cq}}\right)^{\mathrm{T}}, \varepsilon_{\mathrm{Yi}} \sim \mathrm{N}\left(0, \sigma_{\mathrm{e}}^2\right),$$

assume there's no interaction between $\mathrm{A_i}$ and $\mathbf{M_i}$, $\boldsymbol{C_i}$ *denotes covariates*

In the outcome model, the covariates include gender, age (at the measurement of physical activity), BMI, the first 10 genetic principal components and the time difference between measurement of physical activity and measurement of brain imaging. Adding the time difference as a covariate aim to eliminate the impact of the time gap between physical activity and brain imaging measurements.

2) Mediator model:

$$\mathbf{M_i} = \mathrm{A}_1\boldsymbol{\alpha_a} + \boldsymbol{\alpha_c}\mathbf{C_i} + \varepsilon_{\mathrm{M_i}}$$
$$\alpha_{\mathrm{a}} = \left((\alpha_{\mathrm{a}})_1, \ldots, (\alpha_{\mathrm{a}})_{\mathrm{p}}\right)^{\mathrm{T}}, \alpha_{\mathrm{c}} = \left(\alpha_{\mathrm{c1}}^{\mathrm{T}}, \ldots, \alpha_{\mathrm{cq}}^{\mathrm{T}}\right)^{\mathrm{T}},\ \varepsilon_{\mathrm{Mi}} \sim \mathrm{MVN}\left(0, \sum\ \right)$$
assume $\varepsilon_{\mathrm{Yi}}$ and $\varepsilon_{\mathrm{Mi}}$ are independent of $\mathrm{A_i}$, $\mathbf{C_i}$ and each other

In the mediator model, the covariates include gender, age (at the measurement of physical activity), BMI, the first 10 genetic principal components.

As a high-dimensional mediation analysis method, BAMA has two fundamental assumptions:

1) All the mediators contribute small, non-zero effects in mediating the exposure-outcome relationship
2) Only a small proportion of mediators exhibiting additional/large effects.

Based on these two assumptions, normal mixture prior on the coefficients are set in these two models:

$$\left(\beta_{\mathrm{m}}\right)_{\mathrm{j}} \sim \pi_{\mathrm{m}}\mathrm{N}\left(0, \sigma^2_{\mathrm{m1}}\right) + (1-\pi_{\mathrm{m}})\mathrm{N}\left(0, \sigma^2_{\mathrm{m0}}\right),\ \sigma^2_{\mathrm{m1}} > \sigma^2_{\mathrm{m0}}$$

$$(\alpha_{\mathrm{a}})_{\mathrm{j}} \sim \pi_{\mathrm{a}}\mathrm{N}\left(0, \sigma^2_{\mathrm{ma1}}\right) + (1-\pi_{\mathrm{a}})\mathrm{N}\left(0, \sigma^2_{\mathrm{ma0}}\right),\ \sigma^2_{\mathrm{ma1}} > \sigma^2_{\mathrm{ma0}}$$

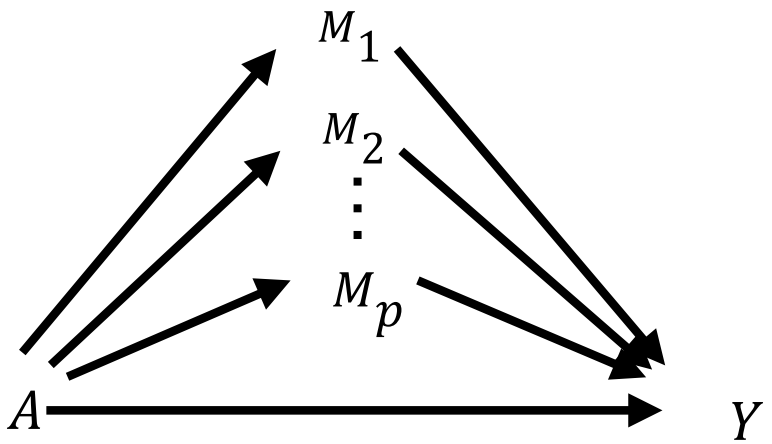

Fig. 5 High-dimensional mediation analysis graph, where A denotes exposure (genetic variants), Y denotes the outcome (SPARE-AD index). $\mathrm{M_i}$ denotes mediator i (certain physical activity pattern). In our study, there are 154 mediators, so p=154

Using a posterior sampling algorithm, we calculate the Posterior Inclusion Probability (PIP), which indicates whether the mediators are active. We introduce indicator variables $\boldsymbol{r_m}, \boldsymbol{r_a} \in \{0,1\}^p$ to indicate which normal component $(\boldsymbol{\beta}_{\boldsymbol{m}})_j$ and $(\boldsymbol{\alpha}_{\boldsymbol{a}})_j$ are from. For mediator j, $r_{mj} = I\left((\boldsymbol{\beta}_{\mathbf{m}}) \sim N\left(0, \sigma^2_{m1}\right)\right), r_{aj} = I\left((\boldsymbol{\alpha}_{\boldsymbol{a}})_j \sim N\left(0, \sigma^2_{ma1}\right)\right)$, where $I(\cdot)$ represents an indicator function. We can then estimate the posterior probability of both $(\boldsymbol{\beta}_{\boldsymbol{m}})_j$ and $(\boldsymbol{\alpha}_{\boldsymbol{a}})_j$ being in the normal components with larger variance as the PIP, defined as $P(r_{mj} = 1, r_{aj} = 1|Data)$. The original paper suggests a more stringent threshold of median inclusion probability of 0.5. Considering the exploratory nature of our analysis, we selected the mediators with PIP>0 as potentially active mediators, and the larger the PIP, the higher the likelihood that the mediator is active. In our analysis, we examine whether each physical activity pattern mediates the effect of each SNP on the SPARE-AD index.

The analysis was conducted using the 'bama' package in statistical software R with the default parameters and a random seed set by us; see below for the command:

$$\text{bama(Y = Y1, A = X1, M = as.matrix(M1), C1 = C11, C2 = C22, method =}$$
$$\text{''BSLMM'', seed = 1245, burnin = 2000, ndraws = 2200, weights = NULL, inits =}$$
$$\text{NULL, control = list (k = 2, lm0 = le - 04, lm1 = 1, lma1 =, l = 1))}$$

This approach used Hastings-within-Gibbs algorithm to obtain posterior samples, and the results are from a Markov chain Monte Carlo (MCMC) approach. The number of iterations to run MCMC before sampling were set by 'burnin' and the default value 2000 was used. 'ndraws' describes the number of draws to take from MCMC that includes burnin draws. We applied the same method as described in Song et al. in 2020 [24, 53] to calculate PIPs for our data.

Univariate mediation analysis

We employed the 'mediate' function from the 'mediation' package [54–58] to conduct univariate mediation analysis. We use bootstrapping with 1000 simulation times for each signal detected from high-dimensional mediation analysis and this analysis aimed to examine the signals identified from high-dimensional mediation analysis, allowing us to compare the performance of univariate mediation analysis with that of high-dimensional mediation analysis.

Results

Genetic variants significantly associated with SPARE-AD index

From the GWAS using SAIGE with the SPARE-AD index as the outcome, we identified a total of 22 SNPs which are significantly associated with SPARE-AD index, which will be used as exposure variables in subsequent analyses. Notably, 20 of these SNPs are located on the *AMPD3* gene. Table 3 shows these 22 SNPs.

Mediation effects identified from high-dimensional mediation analysis

We used 22 SNPs as exposures, 154 physical activity features as mediators, and the SPARE-AD index as the outcome variable. By BAMA, we identified a total of 259 signals whose PIP is greater than 0. Since BAMA does not check whether the relationship

between exposure and mediator is significant, which matters in identifying mediation effects, we filtered out signals where the exposure-mediator relationship was not significant using linear regression (p-value > 0.05) and there are 23 signals remaining. These signals meet the definition of a mediation effect: the mediator is active, as determined by BAMA, the outcome-exposure relationship is significant as confirmed by linear regression, and the outcome-exposure relationship is significant as confirmed by GWAS with the SPARE-AD index as the outcome variable. Table 4 shows these findings.

From Table 4, the patterns of PA always serve as a positive mediator in the signals we identified. The proportion of the mediation effect ranged between 10% and 25% mostly.

Comparison with univariate mediation analysis

The results of univariate mediation analysis for the signals detected from high-dimensional mediation analysis are presented below, allowing for a comparison with the high-dimensional mediation analysis results. Table 5 shows Average Causal Mediation Effect (ACME), which quantifies the proportion of the total effect of the exposure on the outcome that is mediated through the mediator and the corresponding p-value for each signal.

Table 3 List of SNPs significantly associated with the SPARE-AD index. rsID is a unique identifier for a specific SNP. CHR refers to the chromosome on which the SNP is located, and BP stands for the base pair position of the SNP on the chromosome. REF represents the reference allele, and EA stands for effect allele. *P*-value indicates the statistical significance of the association between the SNP and the SPARE-AD index. Gene is the name of the gene in which the SNP is located

rsID	CHR: BP	REF/EA	*P*-value	Gene
rs75589364	chr6:33373881	A/G	1.33E-08	*KIFC1*
rs75086869	chr6:33397796	G/A	3.34E-08	*OMIM*
rs11042786	chr11:10432928	G/C	4.28E-08	*AMPD3*
rs67457110	chr11:10433888	C/T	4.75E-08	*AMPD3*
rs1376001	chr11:10452442	C/T	1.31E-09	*AMPD3*
rs11605232	chr11:10468501	T/C	4.33E-08	*AMPD3*
rs11604780	chr11:10468671	A/G	4.79E-08	*AMPD3*
rs11604833	chr11:10468906	A/G	3.82E-08	*AMPD3*
rs11604838	chr11:10468953	A/G	4.19E-08	*AMPD3*
rs7938316	chr11:10472064	C/T	4.87E-08	*AMPD3*
rs7949917	chr11:10475370	G/A	3.16E-10	*AMPD3*
rs7950039	chr11:10475467	G/A	3.28E-10	*AMPD3*
rs4909930	chr11:10475761	G/A	3.32E-10	*AMPD3*
rs4909931	chr11:10475772	G/T	3.65E-10	*AMPD3*
rs4909932	chr11:10475967	A/G	4.39E-10	*AMPD3*
rs899013	chr11:10476689	A/G	4.62E-10	*AMPD3*
rs7130140	chr11:10489826	A/C	1.77E-09	*AMPD3*
NA	chr11:10498470	T/TAGCA	3.71E-10	*AMPD3*
rs4910143	chr11:10499103	A/G	2.26E-10	*AMPD3*
rs4910144	chr11:10501304	C/G	4.16E-10	*AMPD3*
rs10840425	chr11:10504640	A/C	1.75E-09	*AMPD3*
rs10770119	chr11:10506255	G/T	5.05E-09	*AMPD3*

From Table 5, we could see that all of the 23 signals cannot be detected by univariate mediation analysis at 5% significance level (the p-value for each signal is greater than 0.05).

Discussion

Our results demonstrate that the high-dimensional mediation method allows us to identify more signals than using univariate mediation analysis. This could be due to the consideration of the correlation between the potential mediators derived from complex data objects such as accelerometry data used in our analyses. Our findings are several folds. First, we have identified 3 genetic risk factors that have shown significant associations with both risk of brain atrophy and levels of physical activity. It has been previously reported through a large-scale GWAS study that rs10770119 and rs7949917 are

Table 4 Mediation effects identified from high-dimensional mediation analysis. We identified a total of 23 signals associated with 3 SNPs (rs10770119, rs4909932, rs7949917). We used PIP greater than zero as the criterion to select active mediators; the larger the PIP, the higher the likelihood that the mediator is active. Based on the estimate mediation effect provided by BAMA, we calculated the proportion of the mediation effect by dividing the estimated mediation effect by the total effect (the effect size from the GWAS between the SNP and the SPARE-AD index)

SNP	mediator	PIP	Estimate mediation effect	Proportion
rs10770119	Mean of TLAC	0.011	6.83E-03	25.80%
	Mean of TAC	0.001	2.39E-03	9.03%
	Mean of TLAC (8pm-10pm)	0.001	5.03E-03	18.98%
	Mean of MVPA	0.005	6.58E-03	24.85%
	Mean of M10	0.0005	1.95E-03	7.36%
	SD of TAC	0.002	6.04E-03	22.82%
	SD of MVPA	0.0015	5.02E-03	18.95%
rs4909932	Mean of TAC	0.028	5.80E-03	22.00%
	Mean of TLAC	0.002	2.81E-03	10.67%
	Mean of TLAC (4pm-6pm)	0.0005	2.74E-03	10.42%
	Mean of TLAC (6pm-8pm)	0.0035	4.52E-03	17.15%
	Mean of TLAC (8pm-10pm)	0.001	4.54E-03	17.23%
	Mean of MVPA	0.0435	6.96E-03	26.41%
	SD of TAC	0.0045	5.17E-03	19.62%
	SD of MVPA	0.004	4.74E-03	17.99%
rs7949917	Mean of TAC	0.028	5.80E-03	22.00%
	Mean of TLAC	0.002	2.81E-03	10.67%
	Mean of TLAC (4pm-6pm)	0.0005	2.74E-03	10.42%
	Mean of TLAC (6pm-8pm)	0.0035	4.52E-03	17.15%
	Mean of TLAC (8pm-10pm)	0.001	4.54E-03	17.23%
	Mean of MVPA	0.0435	6.96E-03	26.41%
	SD of TAC	0.0045	5.17E-03	19.62%
	SD of MVPA	0.004	4.74E-03	17.99%

associated with education, cognitive ability, intelligence, and numerical reasoning [59]. Moreover, rs4909932 and rs7949917 were found to be associated with higher white blood cell count and white matter microstructure [60]. Another study showed that a higher white blood cell count is linked to cognitive decline, which implies a higher AD risk [61], confirming our results. There was also a study indicating that white matter disease could be a risk factor for neuronal damage, leading to a higher risk of AD [62]. Our results align with the findings of previous studies, suggesting that these three SNPs could be risk genetic variants for AD. In the meantime, higher levels of physical activity have been shown to be associated with lower white blood cell count [63] and likely decrease the risk of white matter disease by maintaining the white matter microstructure and reducing AD risk [64]. Next, our results identified a positive mediation effect of physical activity patterns where having higher average physical activity intensity levels as quantified using total activity counts (TAC or TLAC), particularly during late afternoon and evening period (during 4pm-10pm), time spent in moderate-to-vigorous activity levels (MVPA) levels of physical activity could compensate part of the risk due to genetic variants. That is, these three SNPs might increase AD risk by decreasing physical activity levels. Although there are studies [28, 30, 32, 65] using mediation analysis to explore the relationships among genetic variants, imaging biomarkers, physical activity, and AD, no current study illustrates the mediation effect of physical activity on the pathway from genetic variants to AD-like brain atrophy index, making our study valuable and insightful. Future interventional studies could be designed with a focus on methods of

Table 5 Average causal mediation effect (ACME), quantifying the average indirect effect

SNP	Mediator	ACME	*p*-value
rs10770119	Mean of TLAC	−8.03E-05	0.422
	Mean of TAC	−9.78E-05	0.34
	Mean of TLAC (8pm-10pm)	−4.20E-05	0.646
	Mean of MVPA	−5.99E-05	0.57
	Mean of M10	−8.24E-06	0.876
	SD of TAC	−1.11E-04	0.292
	SD of MVPA	−8.52E-05	0.466
rs4909932	Mean of TAC	−1.04E-04	0.336
	Mean of TLAC	−9.21E-05	0.324
	Mean of TLAC (4pm-6pm)	−2.20E-05	0.754
	Mean of TLAC (6pm-8pm)	−2.91E-05	0.7
	Mean of TLAC (8pm-10pm)	−3.94E-05	0.644
	Mean of MVPA	−6.94E-05	0.534
	SD of TAC	−9.67E-05	0.28
	SD of MVPA	−8.18E-05	0.45
rs7949917	Mean of TAC	−1.04E-04	0.308
	Mean of TLAC	−9.21E-05	0.352
	Mean of TLAC (4pm-6pm)	−2.20E-05	0.722
	Mean of TLAC (6pm-8pm)	−2.91E-05	0.716
	Mean of TLAC (8pm-10pm)	−3.94E-05	0.66
	Mean of MVPA	−6.94E-05	0.54
	SD of TAC	−9.67E-05	0.306
	SD of MVPA	−8.18E-05	0.442

enhancing daily movement, increasing exercises above certain intensity levels (MVPA) and within certain time of the day (e.g. morning to early afternoon), and assess their potential effects on reducing AD risks.

This study has some limitations. Due to the limited number of features in the dataset used and the difficulty in collecting many features, many covariates highly related to physical activity were not considered. For example, factors such as economic burden [66], built environment [67], crime rates [26], and occupation [68] could significantly influence physical activity levels. Not considering such covariates makes the patterns of physical activity we extracted less accurate, which in turn affects our results. Moreover, due to the scarcity of datasets collecting physical activity and the variation in instruments used to collect physical activity data across different studies and even different stages within the same study, replication datasets are difficult to obtain. Since a significant portion of the participants in UKBB are white British, the lack of replication may make our conclusions less applicable to other populations. Therefore, our study lacks validation from replication datasets, reducing the credibility of the results. In the future, we may utilize larger databases with more information to study the relationship among genetic variants, physical activity, and AD-like brain atrophy index. In terms of the analysis method, BAMA has several drawbacks, and the biggest problem is the assumptions of the model (linear assumption, independence assumption, temporal order assumption, etc.) may be too strong, making it difficult to achieve even with covariates controlled in real-world situations.

Conclusion

Through integrative analysis of multi-omics data, we have identified the mediation pathway of physical activity between genetic factors and AD risk. Overall, genetic factors can increase the brain atrophy measures connected to Alzheimer's disease by reducing physical activity, which may help us better understand the mechanisms of AD cases and provide insights for reducing AD risk and slowing brain aging. Moreover, our research further demonstrates the potential of the high-dimensional mediation analysis method in revealing the mechanisms of disease.

Abbreviations

AD	Alzheimer's disease
ADRD	Alzheimer's disease and related dementias
APOE gene	Apolipoprotein E (APOE) gene
BAMA	Bayesian mediation analysis
BMI	Body mass index
FDA	Functional data analysis
FPCA	Functional principal component analysis
GM	Grey matter
PA	Physical activity
PC	Principal component
PCA	Principal component analysis
PIP	Posterior inclusion probability
SAIGE	Scalable and Accurate Implementation of GEneralized mixed model
SNP	Single nucleotide polymorphism
SPARE-AD index	Spatial Pattern of Abnormality for Recognition of Early Alzheimer's disease index
UKBB	UK Biobank

Acknowledgements

This research has been conducted using data from UK Biobank.

Authors' contributions
Hanxiang Xu: Conceptualization, Methodology, Software, Data analysis, Writing - Original Draft. Shizhuo Mu: Methodology, Software, Data analysis, Writing - Review & Editing. Jingxuan Bao: Data analysis, Writing - Review & Editing. Christos Davatzikos: Data request, Computing Platform, Writing - Review & Editing. Haochang Shou: Supervision, Conceptualization, Methodology, Writing - Review & Editing. Li Shen: Supervision, Conceptualization, Methodology, Writing - Review & Editing.

Funding
This work was supported in part by NIH RF1 AG054409, U01 AG068057, U01 AG066833 and S10OD023495.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Ethics approval is not required. The data used for the preparation of this article is from the UK Biobank (UKBB). The UKBB study has ethical approval, and the ethics committee is detailed here: https://www.ukbiobank.ac.uk/learn-more-about-ukbiobank/governance/ethics-advisory-committee.

Consent for publication
All the authors have read and approved the final version of the manuscript.

Competing interests
We would like to declare a competing interest, as the last author Li Shen is one of the Guest Editors of the Collection.

Received: 31 August 2024 Accepted: 7 February 2025
Published online: 24 March 2025

References

1. Recent research into nicotinamide mononucleotide and ageing. Available from: https://www.nature.com/articles/d42473-022-00002-7. Cited 2024 Mar 28.
2. Eisenstein M. Rejuvenation by controlled reprogramming is the latest gambit in anti-aging. Nat Biotechnol. 2022;40(2):144–6.
3. Silva MVF, Loures CDMG, Alves LCV, de Souza LC, Borges KBG, Carvalho MDG. Alzheimer's disease: risk factors and potentially protective measures. J Biomed Sci. 2019;26(1):33.
4. Yan LL, Li C, Zou S, Li Y, Gong E, He Z, et al. Healthy eating and all-cause mortality among Chinese aged 80 years or older. Int J Behav Nutr Phys Act. 2022;19:60.
5. Alzheimer's disease facts and figures. Alzheimers Dement. 2023;19(4):1598–695.
6. Kuo YC, Rajesh R. Challenges in the treatment of Alzheimer's disease: recent progress and treatment strategies of pharmaceuticals targeting notable pathological factors. Expert Rev Neurother. 2019;19(7):623–52.
7. Cao J, Hou J, Ping J, Cai D. Advances in developing novel therapeutic strategies for Alzheimer's disease. Mol Neurodegener. 2018;13(1):64.
8. d'Errico P, Meyer-Luehmann M. Mechanisms of pathogenic Tau and Aβ protein spreading in Alzheimer's disease. Front Aging Neurosci. 2020;12. Available from: https://www.frontiersin.org/articles/10.3389/fnagi.2020.00265. Cited 2024 Feb 21.
9. Bao J, Chang C, Zhang Q, Saykin AJ, Shen L, Long Q, et al. Integrative analysis of multi-omics and imaging data with incorporation of biological information via structural bayesian factor analysis. Brief Bioinform. 2023;24(2):bbad073.
10. Klimentidis YC, Raichlen DA, Bea J, Garcia DO, Wineinger NE, Mandarino LJ, et al. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. Int J Obes 2005. 2018;42(6):1161–76.
11. Cámara-Calmaestra R, Martínez-Amat A, Aibar-Almazán A, Hita-Contreras F, de Miguel Hernando N, Achalandabaso-Ochoa A. Effectiveness of physical exercise on Alzheimer's disease. Syst Rev J Prev Alzheimers Dis. 2022;9(4):601–16.
12. Stephen R, Hongisto K, Solomon A, Lönnroos E. Physical activity and Alzheimer's disease: a systematic review. J Gerontol Biol Sci Med Sci. 2017;72(6):733–9.
13. Nuzum H, Stickel A, Corona M, Zeller M, Melrose RJ, Wilkins SS. Potential benefits of physical activity in MCI and dementia. Behav Neurol. 2020;2020:7807856.
14. Beinlich FRM, Asiminas A, Untiet V, Bojarowska Z, Plá V, Sigurdsson B, et al. Oxygen imaging of hypoxic pockets in the mouse cerebral cortex. Science. 2024;383(6690):1471–8.
15. Rijnhart JJM, Lamp SJ, Valente MJ, MacKinnon DP, Twisk JWR, Heymans MW. Mediation analysis methods used in observational research: a scoping review and recommendations. BMC Med Res Methodol. 2021;21(1):226.
16. Liu H, Jin IH, Zhang Z, Yuan Y. Social network mediation analysis: a Latent space approach. Psychometrika. 2021;86(1):272–98.
17. Liu J, Ulrich C. Mediation analysis in nursing research: a methodological review. Contemp Nurse. 2016;52(6):643–56.
18. Smyth HL, Pitpitan EV, MacKinnon DP, Booth RE. Assessing potential outcomes mediation in HIV interventions. AIDS Behav. 2021;25(8):2441–54.
19. Klumparendt A, Nelson J, Barenbrügge J, Ehring T. Associations between childhood maltreatment and adult depression: a mediation analysis. BMC Psychiatry. 2019;19(1):36.

20. Mu S, Bao J, Xu H, Shivakumar M, Yang S, Ning X, Kim D, Davatzikos C, Shou H, Shen L. Multivariate mediation analysis with voxel-based morphometry revealed the neurodegeneration pathways from genetic variants to Alzheimer's disease. AMIA Inform Summit. 2024;2024:344–53.
21. Employing Informatics strategies in Alzheimer's disease research: a review from genetics, multiomics, and biomarkers to clinical outcomes | Annual reviews. Available from: https://www.annualreviews.org/content/journals/10.1146/annurev-biodatasci-102423-121021. Cited 2024 Aug 29.
22. Estimation and inference for the indirect effect in high-dimensional linear mediation models | Biometrika | Oxford Academic. Available from: https://academic.oup.com/biomet/article/107/3/573/5829472. Cited 2024 Feb 21.
23. Gao Y, Yang H, Fang R, Zhang Y, Goode EL, Cui Y. Testing mediation effects in high-dimensional epigenetic studies. Front Genet. 2019;10:1195.
24. Song Y, Zhou X, Zhang M, Zhao W, Liu Y, Kardia SLR, et al. Bayesian shrinkage estimation of high dimensional causal mediation effects in omics studies. Biometrics. 2020;76(3):700–10.
25. Chén OY, Crainiceanu C, Ogburn EL, Caffo BS, Wager TD, Lindquist MA. High-dimensional multivariate mediation with application to neuroimaging data. Biostat Oxf Engl. 2018;19(2):121–36.
26. Seefeldt V, Malina RM, Clark MA. Factors affecting levels of physical activity in adults. Sports Med. 2002;32(3):143–68.
27. Chen F, Hu W, Cai J, Chen S, Liu W. Instrumental variable-based high-dimensional mediation analysis with unmeasured confounders for survival data in the observational epigenetic study. Front Genet. 2023;14. Available from: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2023.1092489/full. Cited 2024 Mar 7.
28. Yingxuan E, Yao X, Liu K, Risacher SL, Saykin AJ, Long Q, et al. Polygenic mediation analysis of Alzheimer's disease implicated intermediate amyloid imaging phenotypes. AMIA Annu Symp Proc. 2021;2020:422–31.
29. Jiang J, Hong Y, Li W, Wang A, Jiang S, Jiang T, et al. Chain mediation analysis of the effects of nutrition and cognition on the association of apolipoprotein E ε4 with neuropsychiatric symptoms in Alzheimer's disease. J Alzheimers Dis JAD. 2023;96(2):669–81.
30. Pala D, Lee B, Ning X, Kim D, Shen L. Mediation analysis and mixed-effects models for the identification of stage-specific imaging genetics patterns in Alzheimer's disease. Proc IEEE Int Conf Bioinforma Biomed. 2022;2022:2667–73.
31. Romeo AV, Edney SM, Plotnikoff RC, Olds T, Vandelanotte C, Ryan J, et al. Examining social-cognitive theory constructs as mediators of behaviour change in the active team smartphone physical activity program: a mediation analysis. BMC Public Health. 2021;21(1):88.
32. Chen T, Mandal A, Zhu H, Liu R. Imaging genetic based mediation analysis for human cognition. Front Neurosci. 2022;16. Available from: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2022.824069/full. Cited 2024 Mar 7.
33. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–9.
34. Leroux A, Di J, Smirnova E, Mcguffey EJ, Cao Q, Bayatmokhtari E, et al. Organizing and analyzing the activity data in NHANES. Stat Biosci. 2019;11(2):262–87.
35. Leroux A, Xu S, Kundu P, Muschelli J, Smirnova E, Chatterjee N, et al. Quantifying the predictive performance of objectively measured physical activity on mortality in the UK Biobank. J Gerontol Biol Sci Med Sci. 2020;76(8):1486–94.
36. Xiao L, Zipunnikov V, Ruppert D, Crainiceanu C. Fast covariance estimation for high-dimensional functional data. Stat Comput. 2016;26(1):409–21.
37. Refund.pdf. Available from: https://cran.r-project.org/web/packages/refund/refund.pdf. Cited 2024 Feb 21.
38. Davatzikos C, Xu F, An Y, Fan Y, Resnick SM. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. Brain. 2009;132(8):2026–35.
39. Habes M, Pomponio R, Shou H, Doshi J, Mamourian E, Erus G, et al. The brain chart of aging: machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans. Alzheimers Dement J Alzheimers Assoc. 2021;17(1):89–102.
40. Wen J, Fu CHY, Tosun D, Veturi Y, Yang Z, Abdulkadir A, et al. Characterizing heterogeneity in neuroimaging, cognition, clinical symptoms, and genetics among patients with late-life depression. JAMA Psychiatry. 2022;79(5):464–74.
41. Doshi J, Erus G, Ou Y, Resnick SM, Gur RC, Gur RE, et al. MUSE: MUlti-atlas region segmentation utilizing ensembles of registration algorithms and parameters, and locally optimal atlas selection. NeuroImage. 2016;127:186–95.
42. Yang Q, Gao S, Lin J, Lyu K, Wu Z, Chen Y, et al. A machine learning-based data mining in medical examination data: a biological features-based biological age prediction model. BMC Bioinformatics. 2022;23(1):411.
43. Holzscheck N, Falckenhayn C, Söhle J, Kristof B, Siegner R, Werner A, et al. Modeling transcriptomic age using knowledge-primed artificial neural networks. Npj Aging Mech Dis. 2021;7(1):1–13.
44. Segar MW, Hall JL, Jhund PS, Powell-Wiley TM, Morris AA, Kao D, et al. Machine learning-based models incorporating social determinants of health vs traditional models for predicting in-hospital mortality in patients with heart failure. JAMA Cardiol. 2022;7(8):844–54.
45. Cohen NM, Lifshitz A, Jaschek R, Rinott E, Balicer R, Shlush LI, et al. Longitudinal machine learning uncouples healthy aging factors from chronic disease risks. Nat Aging. 2024;4(1):129–44.
46. Toledo JB, Da X, Bhatt P, Wolk DA, Arnold SE, Shaw LM, et al. Relationship between plasma analytes and SPARE-AD defined brain atrophy patterns in ADNI. PLoS One. 2013;8(2):e55531.
47. Da X, Toledo JB, Zee J, Wolk DA, Xie SX, Ou Y, et al. Integration and relative value of biomarkers for prediction of MCI to AD progression: spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. Neuroimage Clin. 2013;4:164–73.
48. Habes M, Jacobson AM, Braffett BH, Rashid T, Ryan CM, Shou H, et al. Patterns of regional brain atrophy and brain aging in middle- and older-aged adults with type 1 diabetes. JAMA Netw Open. 2023;6(6):e2316182.
49. Bao J, Wen J, Wen Z, Yang S, Cui Y, Yang Z, et al. Brain-wide genome-wide colocalization study for integrating genetics, transcriptomics and brain morphometry in Alzheimer's disease. Neuroimage. 2023;280:120346.
50. Zhou W, Bi W, Zhao Z, Dey KK, Jagadeesh KA, Karczewski KJ, et al. SAIGE-GENE + improves the efficiency and accuracy of set-based rare variant association tests. Nat Genet. 2022;54(10):1466–9.

51. Zhou W, Zhao Z, Nielsen JB, Fritsche LG, LeFaive J, Gagliano Taliun SA, et al. Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. Nat Genet. 2020;52(6):634–9.
52. Yaseen NR, Barnes CLK, Sun L, Takeda A, Rice JP. Genetics of vegetarianism: a genome-wide association study. PLoS One. 2023;18(10):e0291305.
53. Song Y, Zhou X, Kang J, Aung MT, Zhang M, Zhao W, et al. Bayesian sparse mediation analysis with targeted penalization of natural indirect effects. J R Stat Soc Ser C Appl Stat. 2021;70(5):1391–412.
54. Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. Stat Sci. 2010;25(1):51–71.
55. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods. 2010;15(4):309–34.
56. Imai K, Keele L, Tingley D, Yamamoto T. Causal mediation analysis using R. In: Vinod HD, editor. Advances in social science research using. New York: Springer; 2010. p. 129–54. (Lecture Notes in Statistics).
57. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. J Stat Softw. 2014;59:1–38.
58. Imai K, Keele L, Tingley D, Yamamoto T. Unpacking the black box of causality: learning about causal mechanisms from experimental and observational studies. Am Polit Sci Rev. 2011;105(4):765–89.
59. Zhao B, Zhang J, Ibrahim JG, Luo T, Santelli RC, Li Y, et al. Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). Mol Psychiatry. 2021;26(8):3943–55.
60. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The polygenic and monogenic basis of blood traits and diseases. Cell. 2020;182(5):1214-1231.e11.
61. Li J, Zhang Y, Wang H, Guo Y, Shen X, Li M, et al. Exploring the links among peripheral immunity, biomarkers, cognition, and neuroimaging in Alzheimer's disease. Alzheimers Dement Diagn Assess Dis Monit. 2023;15(4):e12517.
62. Bennett S, Thomas AJ. Depression and dementia: cause, consequence or coincidence? Maturitas. 2014;79(2):184–90.
63. Pitsavos C, Chrysohoou C, Panagiotakos DB, Skoumas J, Zeimbekis A, Kokkinos P, et al. Association of leisure-time physical activity on inflammation markers (C-reactive protein, white cell blood count, serum amyloid A, and fibrinogen) in healthy subjects (from the ATTICA study). Am J Cardiol. 2003;91(3):368–70.
64. Konwar S, Manca R, De Marco M, Soininen H, Venneri A. The effect of physical activity on white matter integrity in aging and prodromal to mild Alzheimer's disease with vascular comorbidity. Front Aging Neurosci. 2023;15:1096798.
65. Lenzen S, Gannon B, Rose C, Norton EC. The relationship between physical activity, cognitive function and health care use: a mediation analysis. Soc Sci Med. 2023;335:116202.
66. Ding D, Kolbe-Alexander T, Nguyen B, Katzmarzyk PT, Pratt M, Lawson KD. The economic burden of physical inactivity: a systematic review and critical appraisal. Br J Sports Med. 2017;51(19):1392–409.
67. Kärmeniemi M, Lankila T, Ikäheimo T, Koivumaa-Honkanen H, Korpelainen R. The built environment as a determinant of physical activity: a systematic review of longitudinal studies and natural experiments. Ann Behav Med. 2018;52(3):239–51.
68. Zotcheva E, Bratsberg B, Strand BH, Jugessur A, Engdahl BL, Bowen C, et al. Trajectories of occupational physical activity and risk of later-life mild cognitive impairment and dementia: the HUNT4 70 + study. Lancet Reg Health Eur. 2023;34. Available from: https://www.thelancet.com/journals/lanepe/article/PIIS2666-7762(23)00140-0/fulltext. Cited 2024 Mar 7.

Publisher's Note