



## OPEN ACCESS

## EDITED BY

Carlos Gómez,  
University of Valladolid, Spain

## REVIEWED BY

Vassily Tsytarev,  
University of Maryland, College Park,  
United States  
Tommaso Piccoli,  
University of Palermo, Italy  
Jun Young Park,  
University of Toronto, Canada

## \*CORRESPONDENCE

Joon-Kyung Seong  
jkseong@korea.ac.kr

## SPECIALTY SECTION

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

RECEIVED 22 July 2022

ACCEPTED 24 August 2022

PUBLISHED 20 September 2022

## CITATION

Kim S-W, Song Y-H, Kim HJ, Noh Y,  
Seo SW, Na DL and Seong J-K (2022)  
Unified framework for brain  
connectivity-based biomarkers in  
neurodegenerative disorders.  
*Front. Neurosci.* 16:975299.  
doi: 10.3389/fnins.2022.975299

## COPYRIGHT

© 2022 Kim, Song, Kim, Noh, Seo, Na  
and Seong. This is an open-access  
article distributed under the terms of  
the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution  
or reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Unified framework for brain connectivity-based biomarkers in neurodegenerative disorders

Sung-Woo Kim<sup>1</sup>, Yeong-Hun Song<sup>2</sup>, Hee Jin Kim<sup>3,4,5</sup>,  
Young Noh<sup>6,7</sup>, Sang Won Seo<sup>3,4,8,9</sup>, Duk L. Na<sup>3,10</sup> and  
Joon-Kyung Seong<sup>2,11,12\*</sup>

<sup>1</sup>Department of Bio-Convergence Engineering, Korea University, Seoul, South Korea, <sup>2</sup>Department of Artificial Intelligence, Korea University, Seoul, South Korea, <sup>3</sup>Department of Neurology, Samsung Medical Center, School of Medicine, Sungkyunkwan University, Seoul, South Korea, <sup>4</sup>Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan University, Seoul, South Korea, <sup>5</sup>Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan University, Seoul, South Korea, <sup>6</sup>Department of Neurology, Gil Medical Center, Gachon University of College of Medicine, Incheon, South Korea, <sup>7</sup>Neuroscience Research Institute, Gachon University, Incheon, South Korea, <sup>8</sup>Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Seoul, South Korea, <sup>9</sup>Alzheimer's Disease Convergence Research Center, Samsung Medical Center, Seoul, South Korea, <sup>10</sup>Neuroscience Center, Samsung Medical Center, Seoul, South Korea, <sup>11</sup>School of Biomedical Engineering, Korea University, Seoul, South Korea, <sup>12</sup>Interdisciplinary Program in Precision Public Health, Korea University, Seoul, South Korea

**Background:** Brain connectivity is useful for deciphering complex brain dynamics controlling interregional communication. Identifying specific brain phenomena based on brain connectivity and quantifying their levels can help explain or diagnose neurodegenerative disorders.

**Objective:** This study aimed to establish a unified framework to identify brain connectivity-based biomarkers associated with disease progression and summarize them into a single numerical value, with consideration for connectivity-specific structural attributes.

**Methods:** This study established a framework that unifies the processes of identifying a brain connectivity-based biomarker and mapping its abnormality level into a single numerical value, called a biomarker abnormality summarized from the identified connectivity (BASIC) score. A connectivity-based biomarker was extracted in the form of a connected component associated with disease progression. BASIC scores were constructed to maximize Kendall's rank correlation with the disease, considering the spatial autocorrelation between adjacent edges. Using functional connectivity networks, we validated the BASIC scores in various scenarios.

**Results:** Our proposed framework was successfully applied to construct connectivity-based biomarker scores associated with disease progression, characterized by two, three, and five stages of Alzheimer's disease, and reflected the continuity of brain alterations as the diseases advanced. The BASIC scores were not only sensitive to disease progression, but also specific to the trajectory of a particular disease. Moreover, this framework can be utilized when disease stages are measured on continuous scales, resulting in a notable prediction performance when applied to the prediction of the disease.

**Conclusion:** Our unified framework provides a method to identify brain connectivity-based biomarkers and continuity-reflecting BASIC scores that are sensitive and specific to disease progression.

#### KEYWORDS

brain connectivity, connectivity-based biomarker, biomarker scores, connected component, Laplacian regularization, Kendall's rank correlation, Alzheimer's disease

## 1. Introduction

The brain is composed of structurally and functionally intertwined neurons. This pattern of interconnected neurons can be disrupted or reinforced by the progression of neurodegenerative disorders (Zhan et al., 2016; Lee et al., 2021b; Schumacher et al., 2021; Caravaglios et al., 2022). The development of various neuroimaging devices, such as magnetic resonance imaging (MRI), electroencephalography, and near-infrared spectroscopy, has allowed for non-invasive extraction of the brain's connectivity. The introduction of graph-theoretical methods has allowed for quantitative analysis of the complex structure of the brain's connectivity. Consequently, altered brain connectivity can be quantified, leading to the discovery of connectivity-based biomarkers.

Measurability, particularly the ability to be measured on a continuous scale, is a critical characteristic of diagnostic biomarkers, given their common objectives of explanation and prediction (Shmueli, 2010; Bang, 2020). This need is even more pronounced in neurodegenerative disorders, where multiple pathological phenomena interact in a complex manner (Gomez-Ramirez and Wu, 2014; Jack et al., 2018), complicating attempts to make a diagnosis using a single biomarker. Thus, constructing a single abnormality score for a biomarker can be effectively integrated with other clinical manifestations, such as elucidating the mechanism of the disorder through structural equation modeling or predicting whether a person has the disease through logistic modeling. Furthermore, expressing biomarker abnormalities as a single numerical value assists physicians and patients in intuitively comprehending patients' disease states.

Predictive models, such as linear or logistic regression models, can be used to define biomarker abnormalities, which can be summarized as model responses indicating the level of abnormality (Shen et al., 2017; Lee et al., 2018, 2021a). These models aim to map multiple variables constituting the biomarker to a target variable, either the presence/absence of diseases or a proxy measure representing disease stages. Accurately predicting the target variable in model development may restrict the association between disease stage and response. For example, minimizing the mean squared error (MSE) in

linear predictive models maximizes the association but is limited to a linear relationship. If a disease is represented by a dichotomous variable, the summarized score can be constructed by minimizing the binary cross-entropy (BCE). Nevertheless, this score should ideally be either zero or one, which is incompatible with the fact that the abnormality level reflects continuously changing disease stages. It is also challenging to construct a single model response associated with disease progression when the disease is classified as having three or more consecutive stages.

The structural attributes inherent in biomarkers should also be considered when dealing with them. One attribute of neuroimaging biomarkers is derived from the fact that brain changes can occur over a continuous domain, causing interconnected brain changes in contiguous regions (Lindquist, 2008). A spatially positive correlation has been accounted for by representing disease-related alterations as cluster-like structures comprising adjacent voxels (Poline et al., 1997; Worsley, 2001). This concept has been extended from physical to topological cluster in connectivity-based analysis, where the alteration is mathematically modeled as a graph, and the cluster-like structure is defined as a connected component composed of multiple adjacent edges (Zalesky et al., 2010, 2012; Han et al., 2013). However, spatially autocorrelation should also be considered between the levels of alteration of the edges that comprise the connected component.

This study aimed to establish a framework that unifies the processes of identifying brain connectivity-based biomarkers and mapping abnormality levels into a single numerical value, termed a biomarker abnormality summarized from the identified connectivity (BASIC) score. The framework consists of two successive steps: extraction of altered brain connectivity related to neurodegenerative disorders, and construction of a scoring function with respect to connectivity-based biomarkers. In contrast to predictive models employing MSE or BCE as the objective for model construction, this scoring function was constructed by maximizing Kendall's rank correlation coefficient (Kendall, 1938) between its value and the disease stage, which can be represented by continuous, or even discrete variables having ordinal scales. Especially when the disease stage is represented as a discrete variable, Kendall's correlation is invariant on the nominal scale for data with

binary or ordinal scale; therefore, the association with the disease stage can be measured so long as the nominal scale is maintained. Neighborhood information in brain connectivity is considered in the form of connected components that explicitly reflect spatially positive correlations. We constructed the BASIC scores associated with the progression of Alzheimer's disease (AD) as an example of neurodegenerative disorder, and validated that the scores are sensitive and specific to the disease trajectory. We also validated that the scores also reflect continuous disease-related alterations, resulting in noticeable prediction performance when used to predict the disease.

## 2. Materials and methods

### 2.1. Dataset information

Two independent datasets were utilized in this study. Participants in Datasets 1 and 2 underwent MRI scans and neuropsychological tests at Samsung Medical Center and Gachon University Gil Medical Center, respectively. [Table 1](#) details the demographic and clinical information for each dataset.

Dataset 1 included 116 cognitively unimpaired (CU) individuals, 198 patients with amnesic mild cognitively impairment (MCI), and 212 patients with AD. Patients with MCI and AD were further divided into two subgroups: 57 patients with early MCI (EMCI) and 141 patients with late MCI (LMCI) for the MCI group; and 173 patients with mild AD (MildAD) and 39 patients with moderate-to-severe AD (ModAD) for the AD group. These subgroups were categorized based on clinical criteria, such as diagnostic criteria for MCI/AD and neuropsychological tests. See the previous study for more information on participant recruitment and subgroup division ([Kim et al., 2021](#)). Dataset 2 consisted of 63 CU individuals and 92 AD patients. Similarly, AD patients in Dataset 2 were identified according to clinical diagnostic criteria, which were described in the previous study in conjunction with the recruitment and inclusion/exclusion criteria ([Lee et al., 2021b](#)).

CU participants were further divided into two subgroups based on their ages: young CU (YC) if their age at recruitment was greater than or equal to 65, and old CU (OC) otherwise. Patients with MCI and AD were also subdivided based on onset age: early-onset MCI (EOMCI) or early-onset AD (EOAD) if symptoms developed before age 65, and late-onset MCI (LOMCI) or late-onset MCI (LOAD) otherwise. According to these criteria, individuals in Dataset 1 were classified into 46 YC, 69 EOMCI, 75 EOAD, 70 OC, 129 LOMCI, and 137 LOAD; individuals in Dataset 2 were classified into 33 YC, 51 EOAD, 32 OC, and 41 LOAD.

### 2.2. MR image acquisition

For Dataset 1, both T1-weighted MRI and resting-state functional MRI (rsfMRI) data were acquired using a 3.0-Tesla Intera Achieva MRI scanner (Philips Medical Systems, Best, the Netherlands) as previously described ([Kim et al., 2016, 2021](#)).

For Dataset 2, the T1-weighted MRI data were acquired using a 3.0-Tesla MRI scanner (Verio, Siemens, Erlangen, Germany) as previously described ([Lee et al., 2021b](#)). The rsfMRI data was acquired for 9 min using gradient-echo echo planar imaging (EPI) with the following parameters: TR = 3,000 ms, TE=30 ms, and voxel size =  $3.4 \times 3.4 \times 3.4$  mm<sup>3</sup>. During the acquisition of rsfMRI data, participants were instructed to keep their eyes open, not move their heads, and refrain from falling asleep.

### 2.3. Image preprocessing

The T1-weighted MRI data were preprocessed according to the standard pipeline in FreeSurfer v5.1 to obtain intracranial volume (ICV). The rsfMRI data were preprocessed using FEAT (FMRI Expert Analysis Tool) and FAST (FMRIB's Automated Segmentation Tool) in FSL (FMRIB's Software Library) v5.0 ([Jenkinson et al., 2012](#)) as described in the previous study ([Kim et al., 2021](#)).

### 2.4. PET image acquisition and quantification

Participants in Dataset 2 also underwent [<sup>18</sup>F]-Flutemetamol positron emission tomography (PET) scans with a Siemens Biograph 6 Truepoint PET/computed tomography (CT) scanner (Siemens, Knoxville, Tennessee, United States) using a list-mode emission acquisition. We computed the global amyloid standardized uptake value ratio (SUVR) from each PET image by averaging AD-related regions, including the prefrontal, superior/inferior parietal, lateral temporal, and anterior/posterior cingulate cortices with pons as the reference region. Amyloid positivity was defined as the global SUVR > 0.62. Detailed information on PET imaging parameters and SUVR computation are demonstrated in the previous study ([Lee et al., 2021b](#)).

### 2.5. Functional connectivity network construction

A whole-brain functional connectivity network (FCN) for the *i*-th participant was mathematically modeled as a finite, simple, undirected, and vertex-labeled graph  $\mathcal{G}_i = (\mathcal{V}, \mathcal{E}_i)$ , where  $\mathcal{V}$  is a set of vertices and  $\mathcal{E}_i \in \mathcal{V} \times \mathcal{V}$  is a set of edges.  $\mathcal{V}$  was

TABLE 1 Demographic information according to the stage of Alzheimer's disease continuum.

	Dataset 1					Dataset 2	
	CU	EMCI	LMCI	MildAD	ModAD	CU	AD
$N$	116	57	141	173	39	63	92
Age, years*	65.86 (8.63)	69.63 (7.86)	70.58 (8.86)	71.18 (9.08)	72.51 (9.59)	66.02 (10.97)	68.26 (10.75)
Sex (female), no. <sup>†</sup>	58	32	82	115	30	29	65
Education level, years*	12.07 (4.88)	10.36 (6.41)	11.22 (4.79)	8.95 (5.54)	8.63 (5.66)	12.29 (4.49)	8.57 (4.69)
ICV, $\ell^*$	1.40 (0.22)	1.37 (0.29)	1.34 (0.22)	1.38 (0.21)	1.39 (0.21)	1.50 (0.15)	1.46 (0.16)
MMSE*	28.73 (1.39)	26.95 (2.66)	25.69 (2.81)	20.40 (4.16)	13.74 (5.16)	28.02 (2.02)	17.11 (6.41)
CDR-SOB*	0.51 (0.39)	1.10 (0.79)	1.40 (0.98)	4.34 (1.88)	10.68 (3.69)	0.00 (0.00)	5.23 (3.23)
$A\beta$ deposition							
Global SUVR*			N/A			0.41 (0.11)	1.00 (0.22)
Positivity ( $A\beta+$ ), no. <sup>†</sup>			N/A			2	88

CU, Cognitively unimpaired; EMCI, Early mild cognitive impairment; LMCI, Late mild cognitive impairment; AD, Alzheimer's disease; MildAD, Mild AD; ModAD, Moderate-to-severe AD; ICV, Intracranial volume; MMSE, Mini-Mental State Examination; CDR-SOB, Clinical Dementia Rating scale-Sum Of Boxes; SUVR, Standardized uptake value ratio.

\*Data are given as the mean with standard deviation in parentheses.

<sup>†</sup>Data are given as the number of category in parentheses.

defined as the volumetric regions of interest (ROIs) determined using the Automated Anatomical Labeling (AAL) brain atlas (Tzourio-Mazoyer et al., 2002). The ROIs consist of 40 cerebral cortical regions and five subcortical regions for each hemisphere, resulting in  $|\mathcal{V}| = 90$ . The presence of an edge  $e \in \mathcal{E}_i$ , with a weight of  $c_e$ , representing its connection strength, was determined by temporal associations between regional blood oxygenation level dependent signals, which were computed by averaging the signals of the voxels included in each ROI. Pearson product-moment correlation coefficient (Pearson's  $r$ ) was used for the association measure, followed by the Fisher  $r$ -to- $z$  transformation. A negative coefficient was set to zero, as used in our previous studies (Kim et al., 2017, 2021), because of its ambiguous biological interpretation (Fox et al., 2009; Murphy et al., 2009; Rubinov and Sporns, 2010; Cao et al., 2014). Though, our framework described below is not limited to this thresholding criterion. For convenience, we considered the nonexistent edge  $e \notin \mathcal{E}_i$  to be equivalent to the edge  $e \in \mathcal{E}_i$  with  $c_e = 0$ .

## 2.6. A unified framework for connectivity-based biomarkers

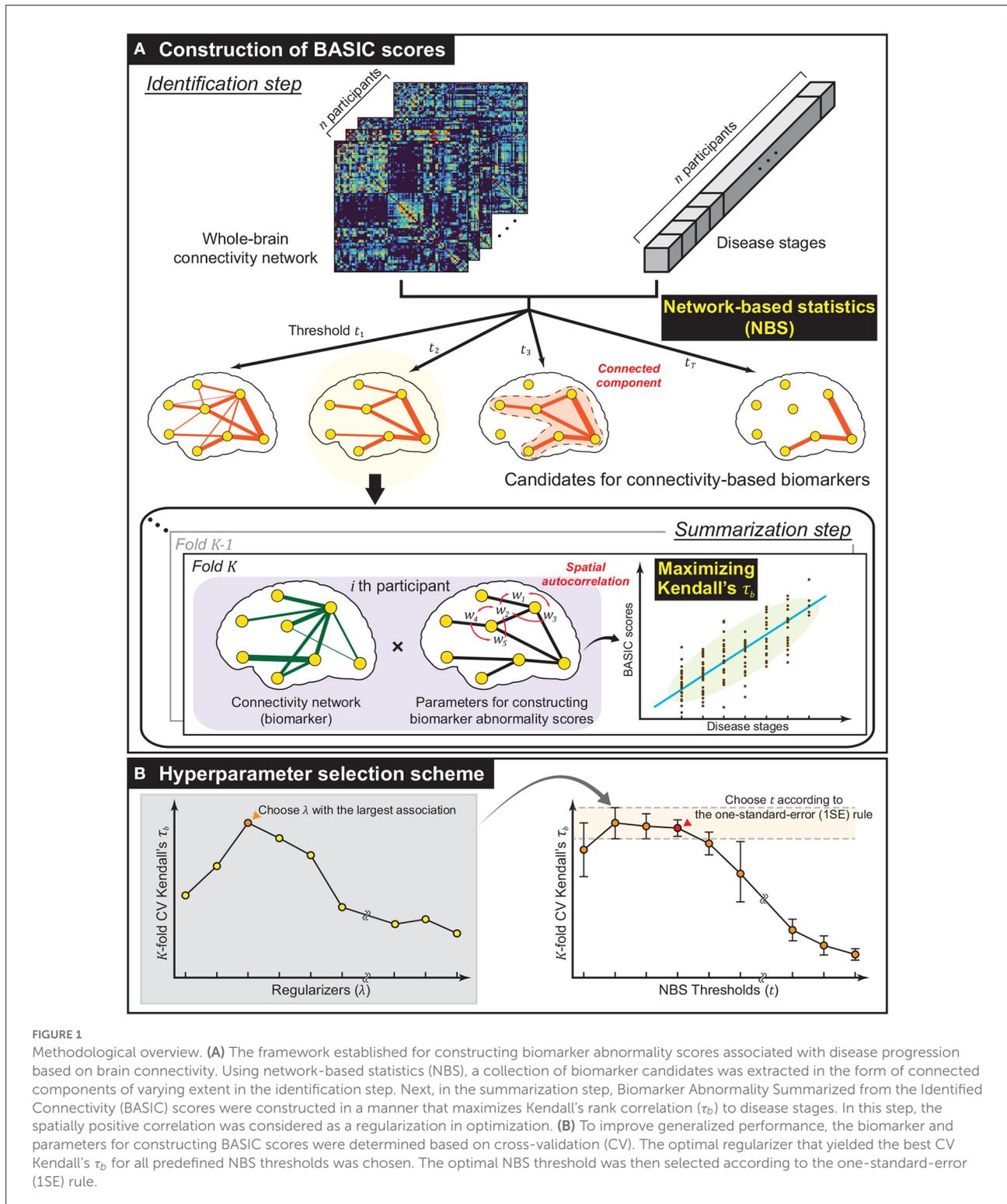
### 2.6.1. Overview

Let  $\mathcal{D} = \{(\mathcal{G}_i, d_i)\}_{i=1}^n$  be a dataset consisting of the whole-brain connectivity network  $\mathcal{G}_i = (\mathcal{V}, \mathcal{E}_i)$ , where  $d_i$  is the realization of a reference variable  $D$  representing the stage of

a disease measured on continuous, dichotomous, or ordinal scales. A *connectivity-based biomarker*, denoted as  $\mathcal{H} = (\mathcal{S}, \mathcal{F})$ , was defined as a connected component of an unweighted graph  $\bar{\mathcal{G}} = (\mathcal{V}, \mathcal{E}_1 \cup \dots \cup \mathcal{E}_n)$ . The abnormality score of  $\mathcal{H}$  was defined as a continuous and bounded real number. More formally, we defined a *scoring function*  $s: \mathcal{G} \rightarrow [0, 1]$ , which produces a *biomarker abnormality score* from  $\mathcal{G}_i[\mathcal{F}] = (\mathcal{V}'_i, \mathcal{E}'_i)$ , an edge-induced subgraph of  $\mathcal{G}_i$  consisting of  $\mathcal{E}'_i \subset \mathcal{E}_i$  with vertices incident to  $e \in \mathcal{E}'_i$ . The framework was designed to identify  $\mathcal{H}$  based on  $\mathcal{D}$ , and obtain each individual's biomarker abnormality score  $s(\mathcal{G}_i[\mathcal{F}])$ , the BASIC score, by finding the optimal  $s$  according to two consecutive steps: *identification* and *summarization*. This process is schematically illustrated in Figure 1A.

### 2.6.2. Identification step

This step extracts a set of possible connectivity-based biomarkers based on  $\mathcal{D}$ . To satisfy the condition that  $\mathcal{H}$  is connected, we employ a network-based statistic (NBS) that identifies the statistically significant connected component for specific hypothesis testing (Zalesky et al., 2010). In the NBS, statistics for the testing, such as  $t$ -statistics, correlation coefficients, and regression coefficients, were calculated on all edges  $e \in \mathcal{E}_1 \cup \dots \cup \mathcal{E}_n$ . Edges whose statistics exceeded a predefined threshold  $t$  were chosen, and the connected components comprising suprathreshold edges were extracted. This procedure was repeated  $m$  times, with the sample being



**FIGURE 1** Methodological overview. **(A)** The framework established for constructing biomarker abnormality scores associated with disease progression based on brain connectivity. Using network-based statistics (NBS), a collection of biomarker candidates was extracted in the form of connected components of varying extent in the identification step. Next, in the summarization step, Biomarker Abnormality Summarized from the Identified Connectivity (BASIC) scores were constructed in a manner that maximizes Kendall's rank correlation ( $\tau_b$ ) to disease stages. In this step, the spatially positive correlation was considered as a regularization in optimization. **(B)** To improve generalized performance, the biomarker and parameters for constructing BASIC scores were determined based on cross-validation (CV). The optimal regularizer that yielded the best CV Kendall's  $\tau_b$  for all predefined NBS thresholds was chosen. The optimal NBS threshold was then selected according to the one-standard-error (1SE) rule.

generated from randomization of the original one, resulting in different statistics for all edges; thus, it yielded distinct connected components with maximum sizes. It gives a null distribution of the maximum extent of connected components,

and statistical significance was determined at the level of connected components rather than at the individual edges.

Edge-level statistics were computed using a general linear model. For each edge  $e \in \mathcal{E}_1 \cup \dots \cup \mathcal{E}_n$ ,  $c_e$  was used as a dependent

variable, and the reference variable  $D$ , together with an intercept and nuisance variables, were used as independent variables. If  $D$  is measured on a continuous scale, then the variable of interest has a single continuous value; otherwise, the variables of interest are made up of dummy variables. Dummy variables were encoded using a staircase coding scheme, particularly when  $D$  was measured on an ordinal scale. To generate randomly permuted samples, we employed the Freedman-Lane procedure (Freedman and Lane, 1983; Winkler et al., 2014), where the permutation was performed on the residual of the reduced model containing only nuisance variables.

### 2.6.3. Summarization step

$\mathcal{H}$  is summarized as a single numerical value in this step. In constructing  $s$ , the following conditions were considered: a) if a sequence of the scores  $\{s(\mathcal{G}_i[\mathcal{F}])\}$  was associated with that of the corresponding disease stage  $\{d_i\}$ , b) if the connectivity-specific structural attribute, a spatially positive correlation on adjacent edges, was explicitly reflected, and c) if  $\inf s = 0$ , and  $\sup s = 1$ . Complying with these conditions, we defined  $s = \Phi_{\mu,\sigma} \circ f$ , where  $f$  is a real-valued function on  $\mathcal{G}$ , called a *summarizing function*, and  $\Phi_{\mu,\sigma}$  is a cumulative distribution function (CDF) of a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ , which are computed as the sample mean and sample standard deviation of  $\{s(\mathcal{G}_i[\mathcal{F}])\}_{i=1}^n$ , respectively.

In this study, we defined  $f(\cdot; \mathbf{w})$ , parameterized by a column vector  $\mathbf{w} \in \mathbb{R}^{|\mathcal{F}|}$ , as the linear combination of the edge connection strengths, that is,

$$f(\mathcal{G}_i[\mathcal{F}]; \mathbf{w}) = \sum_{e \in \mathcal{E}'_i} w_e \cdot (c_e)_i = \mathbf{w}^T \mathbf{c}_i, \tag{1}$$

where  $w_e$  is a parameter of  $f$  for an edge  $e$ , and  $\mathbf{c} \in \mathbb{R}^{|\mathcal{F}|}$  is a column vector whose element is  $c_e$  for an edge  $e$ . Notably, Equation (1) can be applied to graphs with different sets of edges because we set  $c_e = 0$  for even  $e \in \mathcal{F} \setminus \mathcal{E}_i$ .  $\mathbf{w}$  was computed by maximizing the association between the values of  $f$  and the corresponding ones for  $D$ , considering the connectivity-specific spatial autocorrelation simultaneously, or formally satisfying the following objective:

$$\underset{\mathbf{w}}{\text{minimize}} \quad -\mathcal{R}(w^T \mathbf{c}_i, d_i; \mathcal{D}) + \lambda \|\mathbf{B}\mathbf{w}\|_2^2, \tag{2}$$

where  $\mathcal{R}(\cdot, \cdot; \mathcal{D})$  is the measure of association between two univariate variables in the dataset  $\mathcal{D}$ ,  $\mathbf{B} \in \{-1, 0, 1\}^{|\mathcal{S}| \times |\mathcal{F}|}$  is an oriented vertex-edge incidence matrix, and  $\lambda$  is a Tikhonov regularizer. For  $\mathcal{R}$ , Kendall's tau-b ( $\tau_b$ ) was used (Kendall, 1938). In contrast to Pearson's  $r$ ,  $\tau_b$  can be used for the association between two variables measured on a dichotomous or ordinal scale, as well as for identifying nonlinear monotonic relationships beyond linear relationships (Chok, 2010). Further, as a rank statistic,  $\tau_b$  is more resistant to outliers when calculated

for continuous variables (Couso et al., 2018). The second term is for regularization, which can be rewritten as follows:

$$\mathbf{w}^T \mathbf{B}^T \mathbf{B} \mathbf{w} = \sum_{u,v \in \mathcal{F}: u \sim v} (w_u - w_v)^2, \tag{3}$$

where  $u \sim v$  means that the edges  $u$  and  $v$  are adjacent. Note that the  $\mathbf{B}^T \mathbf{B}$  is equal to the edge-based graph Laplacian, and the regularization term results in Laplacian regularization (Ando and Zhang, 2006). In other words, spatial autocorrelation is added in the form of minimizing of the difference between the parameter values of adjacent edges.

We solved the optimization problem with the objective (Equation 2) using the GRID algorithm that was known to find a solution comparable to the globally optimal one (Croux et al., 2007; Alfons et al., 2017). It deals with a slightly modified version of objective (Equation 2), which has an additional equality constraint  $\|\mathbf{w}\|_2 = 1$ , resulting in a new optimization problem to find the optimal direction of  $\mathbf{w}$ . Finally, the final objective is stated as follows:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & -\mathcal{R}(w^T \mathbf{c}_i, d_i; \mathcal{D}) + \lambda \|\mathbf{B}\mathbf{w}\|_2^2, \\ \text{subject to} \quad & \|\mathbf{w}\|_2 = 1. \end{aligned} \tag{4}$$

In the first iteration of this algorithm, the optimal direction was determined by performing a grid search that divides the angular intervals  $[-\pi/2, \pi/2]$  into  $n_g$  equal parts alternately for each coordinate. At  $r$ th iteration, the search interval was reduced to  $[-\pi/2^r, \pi/2^r]$ .

Finally, the BASIC score for a given whole-brain connectivity network  $\mathcal{G}_i$  is calculated as  $\hat{\mathbf{w}}^T \mathbf{c}_i$ , where  $\hat{\mathbf{w}}$  is the optimal solution of the optimization problem (Equation 4).

### 2.6.4. Hyperparameter selection scheme

The BASIC score aims to improve generalized performance in association with disease progression by selecting the value of hyperparameters based on cross-validation (CV). Two hyperparameters were selected for the proposed framework:  $t$  and  $\lambda$ . The former controls the extent of the connectivity-based biomarker, analogous to controlling the sparsity of parameters of  $f$ . The latter controls for the degree of spatial positive correlation. The optimal values of the hyperparameters were determined by simultaneously varying their values within predefined sets.

For a fixed value of  $t$ ,  $K$ -fold CV was performed at each value using a set of predefined values of  $\lambda$ , and the value with the highest averaged  $\tau_b$  across  $K$  folds was selected. Then,  $t$  was chosen such that the corresponding subnetwork exhibits the most parsimonious and comparable average  $\tau_b$  according to the one-standard-error (1SE) rule (Hastie et al., 2009). The scheme for selecting the values of  $t$  and  $\lambda$  is illustrated in Figure 1B.

## 2.7. Experiments

To validate whether the framework performs well when  $D$  is measured on a continuous or ordinal scale, we identified connectivity-based biomarkers related to various stages of AD and A $\beta$  deposition from whole-brain FCNs using the optimal values of  $t$  and  $\lambda$  determined by  $K$ -fold CV. We also computed the corresponding BASIC scores using the entire dataset. Table 2 provides detailed information on the experimental settings. The optimal value of  $t$  was determined by computing the generalized association performance of sub-networks containing 0.1–10% of the total number of possible edges (i.e., from four to 401) for values within its range because the NBS is ineffective and has a low statistical power at too high or too low threshold (Zalesky et al., 2010).

When  $D$  was measured on a dichotomous scale, we compared the BASIC scores with scores determined by the logistic model, where the score function was defined as  $s = g \circ f$  with a sigmoid function  $g(x) = 1/(1 + e^{-x})$ . In contrast to the BASIC score, where  $\mathbf{w}$  satisfies the objective with respect to the summarizing function  $f$ , the logistic model-based scores were obtained by minimizing the BCE between  $s(\mathcal{G}_i[\mathcal{F}]) = g(f(\mathcal{G}_i[\mathcal{F}]; \mathbf{w}))$  and  $d_i \in \{0, 1\}$  as the target variable. For  $D$  with continuous values, we compared the BASIC scores with scores determined by optimizing the least-square form of the classical canonical correlation with Laplacian regularization (Sun et al., 2008).

### 2.7.1. Statistical analysis

Continuous biomarker abnormality scores in two successive groups along a disease continuum were compared using the Student's  $t$ -test and Welch's  $t$ -test. The false discovery rate (FDR) procedure was used for each experiment to correct for multiple comparisons (Benjamini and Hochberg, 1995). To examine the relationship between biomarker abnormalities and cognitive function, we calculated the coefficients of multiple correlations between the biomarker score and clinical scores, including the mini-mental state examination (MMSE) and the clinical dementia rating scale-sum of boxes (CDR-SOB) scores. Statistical significance of the coefficient of multiple correlations was determined by permutation testing with 10,000 permutations.

### 2.7.2. Classification analysis

Support vector machine with radial basis function kernel and uniform class prior probability was used for binary classification. It is run by MATLAB R2021b ([www.mathworks.com](http://www.mathworks.com)), with all other hyperparameters set to default.

## 3. Results

The optimal values of the hyperparameters and corresponding results for all experiments are presented in Table 3.

### 3.1. Continuous disease-related alteration in BASIC scores

As an example of a neurodegenerative disorder, we first computed the BASIC scores associated with the presence or absence of AD for Dataset 1, (*Experiment 1*). By examining the distribution of estimated scores for all CU individuals and AD patients, we found that the likelihood of intermediate scores was comparable to that of extreme CU or AD scores (Figure 2A). This phenomenon differed from the biomarker abnormality scores computed using the logistic model (*Experiment 1-1*), demonstrating a higher likelihood of extreme CU and AD scores and a very low likelihood of the intermediate scores (Figure 2B).

Upon estimating the distributions for CU and AD individuals, the distribution of BASIC scores in each group was completely distinguished, as maximizing Kendall's  $\tau_b$  with a dichotomous variable equivalent to maximizing the area under the receiver operating characteristic (AUROC) curve (Hanley and McNeil, 1982; Brossart et al., 2018). However, the score distributions of the CU and AD groups were differently shaped for each method. In contrast to the logistic model-based scores, we discovered that the BASIC scores were evenly distributed within each score range for each group.

When scores of MCI patients in Dataset 1 were calculated with the parameters of the summarizing function obtained for CU and AD individuals in Dataset 1, their BASIC scores had the highest likelihood in the mid-scores (gray-shaded area in Figure 2A). Patients with MCI were expected to be distributed between CU and AD. In contrast, a higher likelihood was observed in the score range close to AD in the logistic model-based scores (gray-shaded area in Figure 2B). A similar trend was observed when the scores were computed for Dataset 2 (Figures 2C,D). This demonstrates that BASIC scores reflect disease progression along the AD continuum more accurately than logistic model-based scores.

### 3.2. BASIC scores for multiple disease stages

AD is characterized by the gradual progression of cognitive deficits along a continuum, and its progression can be discretized into more than two consecutive stages. The proposed framework is applicable to this situation, producing BASIC scores that are associated with disease progression. Figure 3

TABLE 2 List of experimental settings.

Experiment	1	2	3	4	5
Disease/Pathology	AD	AD	AD	EOAD/LOAD	A $\beta$ deposition
Reference variable ( <i>D</i> )	Dichotomous	Ordinal	Ordinal	Ordinal	Continuous
Subgroups	CU, AD	CU, MCI, AD	CU, EMCI, LMCI, MildAD, ModAD	YC, YOMCI, YOAD / OC, LOMCI, LOAD	
Dataset ( <i>D</i> )	Dataset 1	Dataset 1	Dataset 1	Dataset 1	Dataset 2
Hyperparameters					
<b>Identification step</b>					
Edge-level statistics			partial <i>F</i> -statistics		
No. of permutation ( <i>m</i> )			10,000		
NBS threshold ( <i>t</i> )	{ <i>k</i> } <sub><i>k</i>=1</sub> <sup>30</sup>	{0.5 <i>k</i> } <sub><i>k</i>=1</sub> <sup>30</sup>	{0.5 <i>k</i> } <sub><i>k</i>=1</sub> <sup>40</sup>	{ <i>k</i> } <sub><i>k</i>=1</sub> <sup>30</sup>	{ <i>k</i> } <sub><i>k</i>=1</sub> <sup>20</sup>
Nuisance variables			age, sex, education level, and ICV		
<b>Summarization step</b>					
No. of iterations ( <i>n<sub>r</sub></i> )			100		
No. of grid points ( <i>n<sub>g</sub></i> )			20		
No. of folds ( <i>K</i> )			5		
Regularizer ( $\lambda$ )			{ <i>e</i> <sup>-15</sup> , <i>e</i> <sup>-14</sup> , ..., <i>e</i> <sup>1</sup> , <i>e</i> <sup>0</sup> }		

CU, Cognitively unimpaired; MCI, Mild cognitive impairment; EMCI, Early MCI; LMCI, Late MCI; AD, Alzheimer's disease; MildAD, Mild AD; ModAD, Moderate-to-severe AD; NBS, Network-based statistics; ICV, Intracranial volume.

TABLE 3 Selected hyperparameter values and the corresponding results for each experiment.

	Diseases	Disease stages	Objective functions	Regularizations	<i>t</i>	$\lambda$	No. of edges	5-fold CV performances*
Exp. 1	AD	2 stages	Kendall's $\tau_b$	Laplacian	9	<i>e</i> <sup>-10</sup>	215	0.5493 (0.0308) <sup>†</sup>
Exp. 2	AD	3 stages	Kendall's $\tau_b$	Laplacian	6	<i>e</i> <sup>-4</sup>	160	0.4899 (0.0549) <sup>†</sup>
Exp. 3	AD	5 stages	Kendall's $\tau_b$	Laplacian	4	<i>e</i> <sup>-10</sup>	132	0.3993 (0.0305) <sup>†</sup>
Exp. 4a	EOAD	3 stages	Kendall's $\tau_b$	Laplacian	8	<i>e</i> <sup>-12</sup>	80	0.5427 (0.0889) <sup>†</sup>
Exp. 4b	LOAD	3 stages	Kendall's $\tau_b$	Laplacian	7	<i>e</i> <sup>-12</sup>	39	0.4192 (0.0523) <sup>†</sup>
Exp. 5	A $\beta$	Continuous	Kendall's $\tau_b$	Laplacian	8	<i>e</i> <sup>9</sup>	126	0.4839 (0.0566) <sup>†</sup>
Exp. 1'	AD	2 stages	Kendall's $\tau_b$	None	10		154	0.4824 (0.0319) <sup>†</sup>
Exp. 2'	AD	3 stages	Kendall's $\tau_b$	None	6		160	0.4298 (0.0574) <sup>†</sup>
fExp. 3'	A $\beta$	5 stages	Kendall's $\tau_b$	None	4.5		81	0.3638 (0.0590) <sup>†</sup>
Exp. 1-1	AD	2 stages	BCE	Laplacian	7	<i>e</i> <sup>-12</sup>	360	0.8445 (0.0263) <sup>‡</sup>
Exp. 5-1	A $\beta$	Continuous	Pearson's <i>r</i>	Laplacian	8	<i>e</i> <sup>-16</sup>	126	0.7019 (0.0527) <sup>§</sup>

\*5-fold CV performances are given as the mean and standard deviation.

<sup>†</sup>5-fold CV performances were computed for Kendall's  $\tau_b$ .

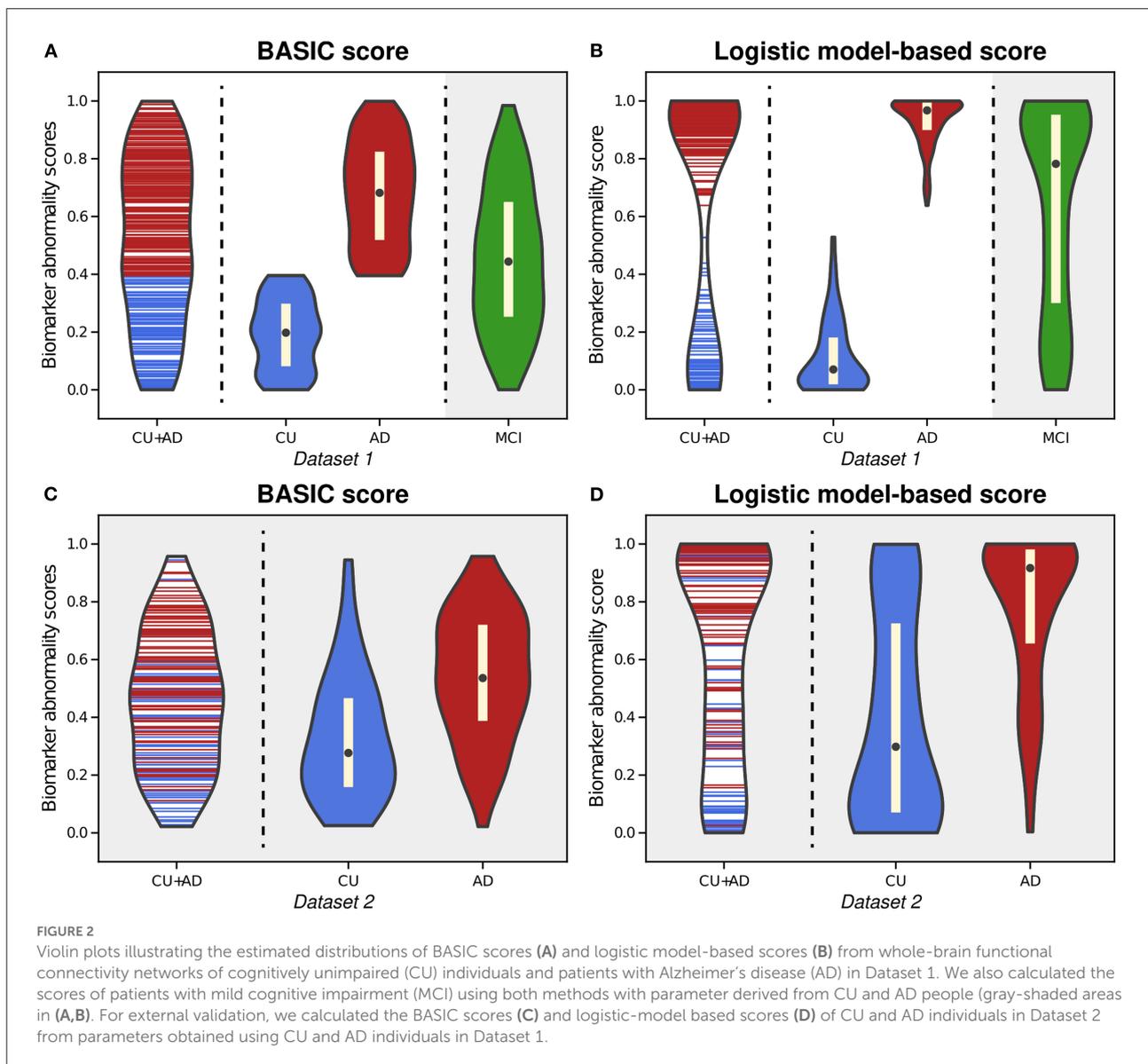
<sup>‡</sup>5-fold CV performances were computed for accuracy.

<sup>§</sup>5-fold CV performances were computed for Pearson's *r*.

CV, Cross-validation; Exp., Experiment; AD, Alzheimer's disease; EOAD, Early-onset AD; LOAD, Late-onset AD; BCE, Binary cross-entropy.

depicts the estimated distributions when the stages on the AD continuum are three (CU, MCI, and AD; *Experiment 2*) and five (CU, EMCI, LMCI, MildAD, and ModAD; *Experiment 3*). In each instance, the distributions of scores for CU individuals and AD patients were more naturally at their boundaries than

when scores were determined using only two groups. The MCI patients' scores exhibited a bell-shaped distribution, with the highest likelihood between the CU and AD scores. When using five groups, the median difference between LMCI and MildAD was larger than that between EMCI and LMCI or between

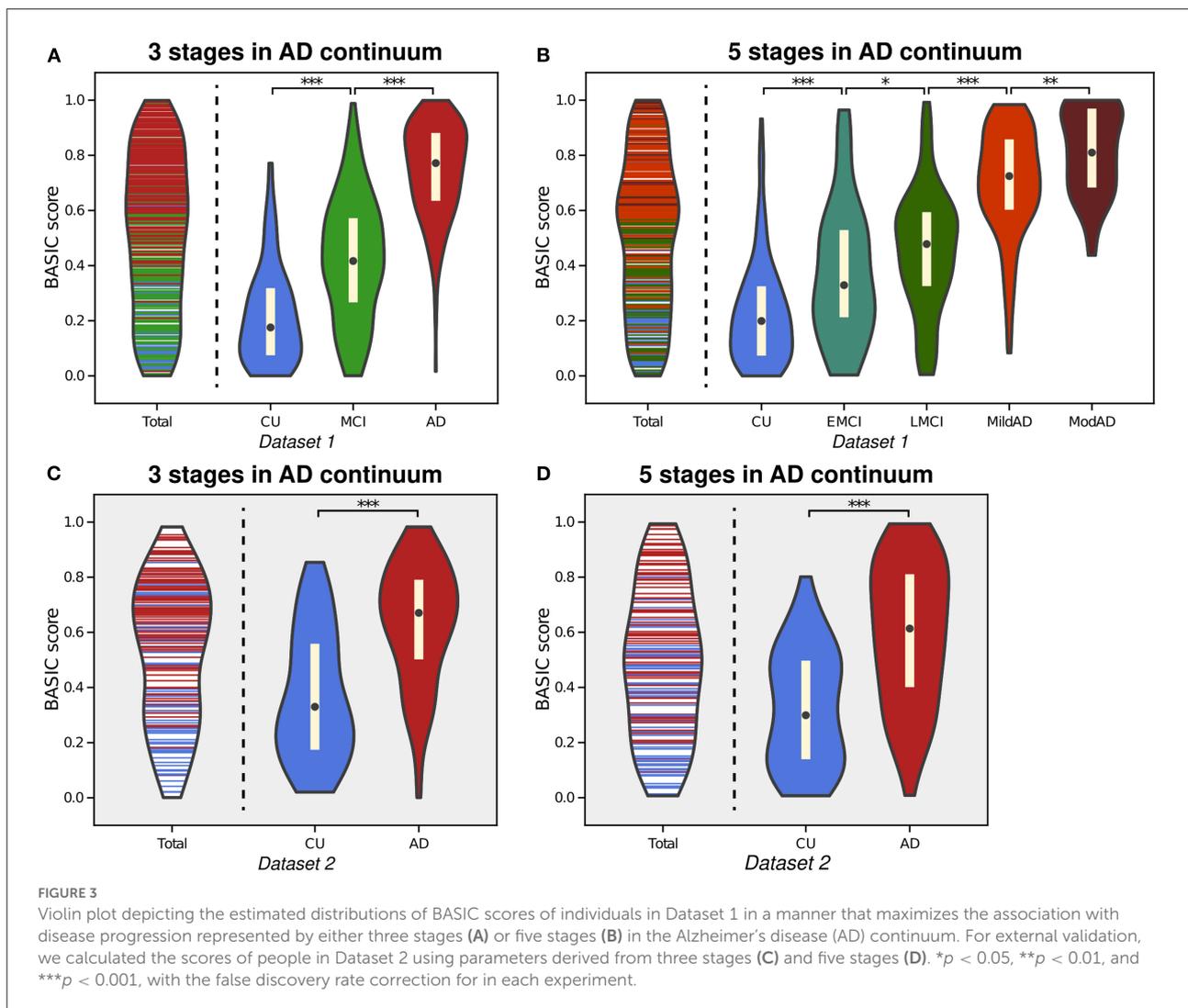


MildAD and ModAD, indicating that BASIC scores can reflect monotonic nonlinear changes across disease stages.

We performed correlation analyses between the BASIC scores determined using different numbers of stages on the AD continuum and cognitive functioning scores directly related to AD (Figure 4). When BASIC scores were calculated using only two stages, there were significant correlations for all participants as well as only patients with dementia (including both MCI and AD patients). However, within each of the MCI and AD groups, no significant correlation was found. All associations were strengthened when disease stages were further subdivided, particularly within the MCI and AD groups, implying that the greater the number of disease stages utilized, the more truthful the biomarker score.

### 3.3. Edge importance in connectivity-based biomarkers

To determine which edge had the greatest impact on biomarker scores, we defined *edge importance* as the square of the parameter values on edges, similar to how feature importance is calculated in linear regression (Grömping, 2015). As a representative example, the profile of edge importance in constructing BASIC scores associated with AD progression represented by the five stages is illustrated in Figure 5. Connections with high edge importance were incident to a set of regions, including the opercular/orbital parts of the left inferior frontal gyrus, left thalamus, right parahippocampal gyrus, right supramarginal gyrus, and right precuneus, where



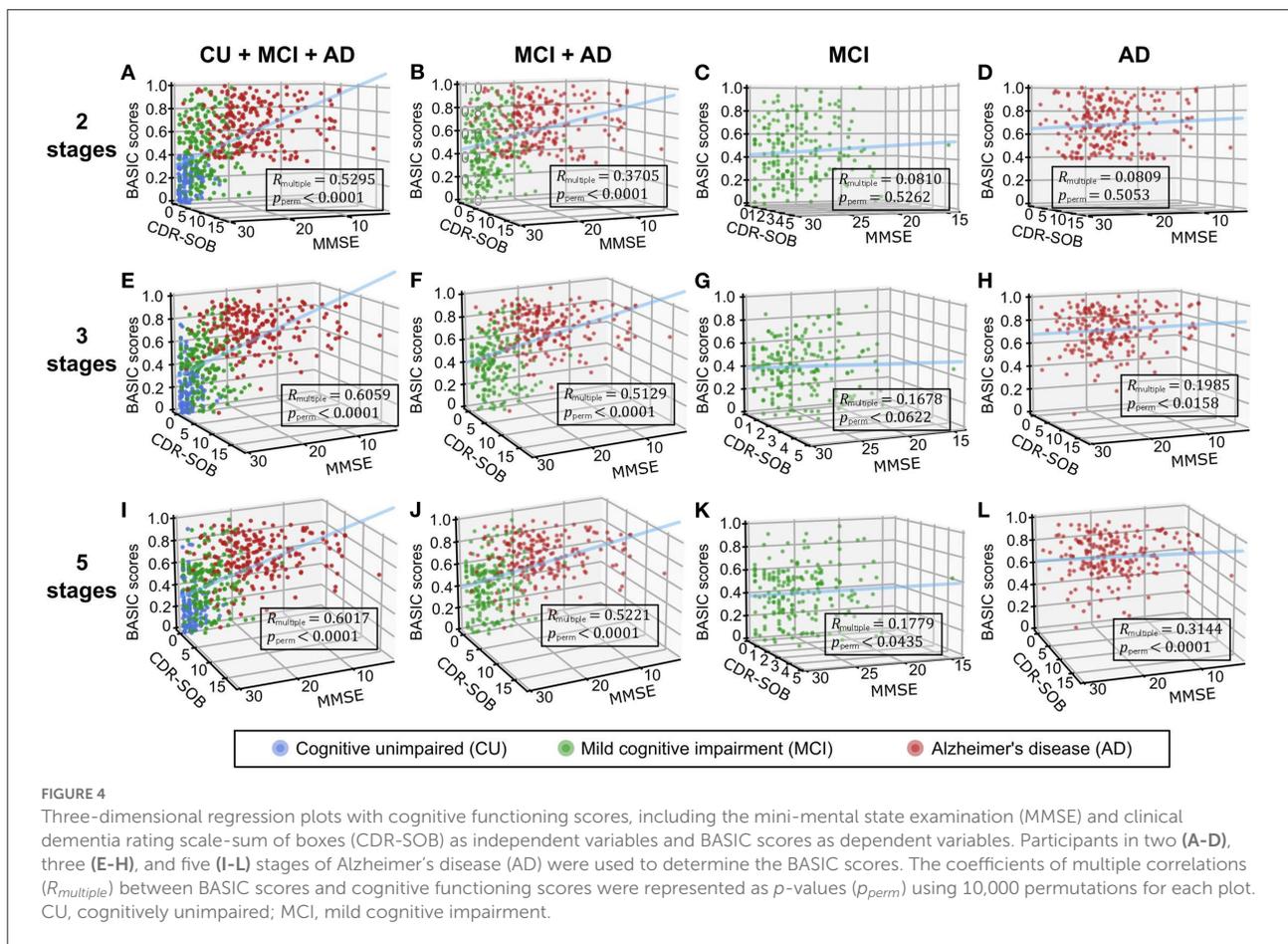
functional connectivity is altered as AD progresses (Grady et al., 2001; Wang et al., 2007; Zhou et al., 2013; Li et al., 2017).

### 3.4. Disease trajectory-specific BASIC scores

AD can be subdivided according to various criteria, one of which is the subtype divided by onset age: EOAD and LOAD. Previous research has shown that EOAD and LOAD can be distinguished by their distinct pathological and clinical characteristics, including patterns of cortical atrophy, white matter disruption, and cognitive impairment (Panegyres and Chen, 2013; Mendez, 2019; Lee et al., 2021b). In this study, our proposed framework provided BASIC scores for EOAD-related biomarkers based on EOAD trajectory (YC-EOMCI-EOAD)

and LOAD-related biomarkers based on LOAD trajectory (OC-LOMCI-LOAD), respectively (*Experiment 4a* and *4b*). BASIC scores for the EOAD-related and LOAD-related biomarkers were sensitive to disease progressions of each subtype (left parts of Figures 6A,B).

Notably, individuals in the LOAD trajectory had relatively high BASIC scores for the EOAD-related biomarkers; in contrast, individuals in the EOAD trajectory had relatively low LOAD-related biomarker scores (gray-shaded areas in Figures 6A,B). Similar tendencies were observed for individuals in Dataset 2 when their BASIC scores were calculated using the parameters estimated from Dataset 1 (Figures 6C,D). This is in line with previous results that found that LOAD has a typical memory-disorder manifestation, whereas EOAD is characterized by deficits in the non-memory domain, such as visuoconstructional skills, executive function, and memory function (Joubert et al., 2016; Tellechea et al., 2018). In other words, the symptoms of EOAD are more extensive than those



of LOAD, and features contributing to the EOAD-related biomarker score may also be present in LOAD, resulting in relatively high EOAD-related biomarker scores for individuals in the LOAD trajectory. Similarly, the LOAD-related biomarker score may be lower in EOAD because LOAD-related scores do not include all characteristics associated with EOAD. It is also worth noting that comparing the score distributions of stages on different disease trajectories, beyond comparing degrees of increases for the trajectories, may not be meaningful, because age or other demographic/clinical variables are not considered when constructing the score. These variables can distort the level of biomarker abnormality due to their association with disease manifestation; their effect can be controlled in further analyses that use the scores.

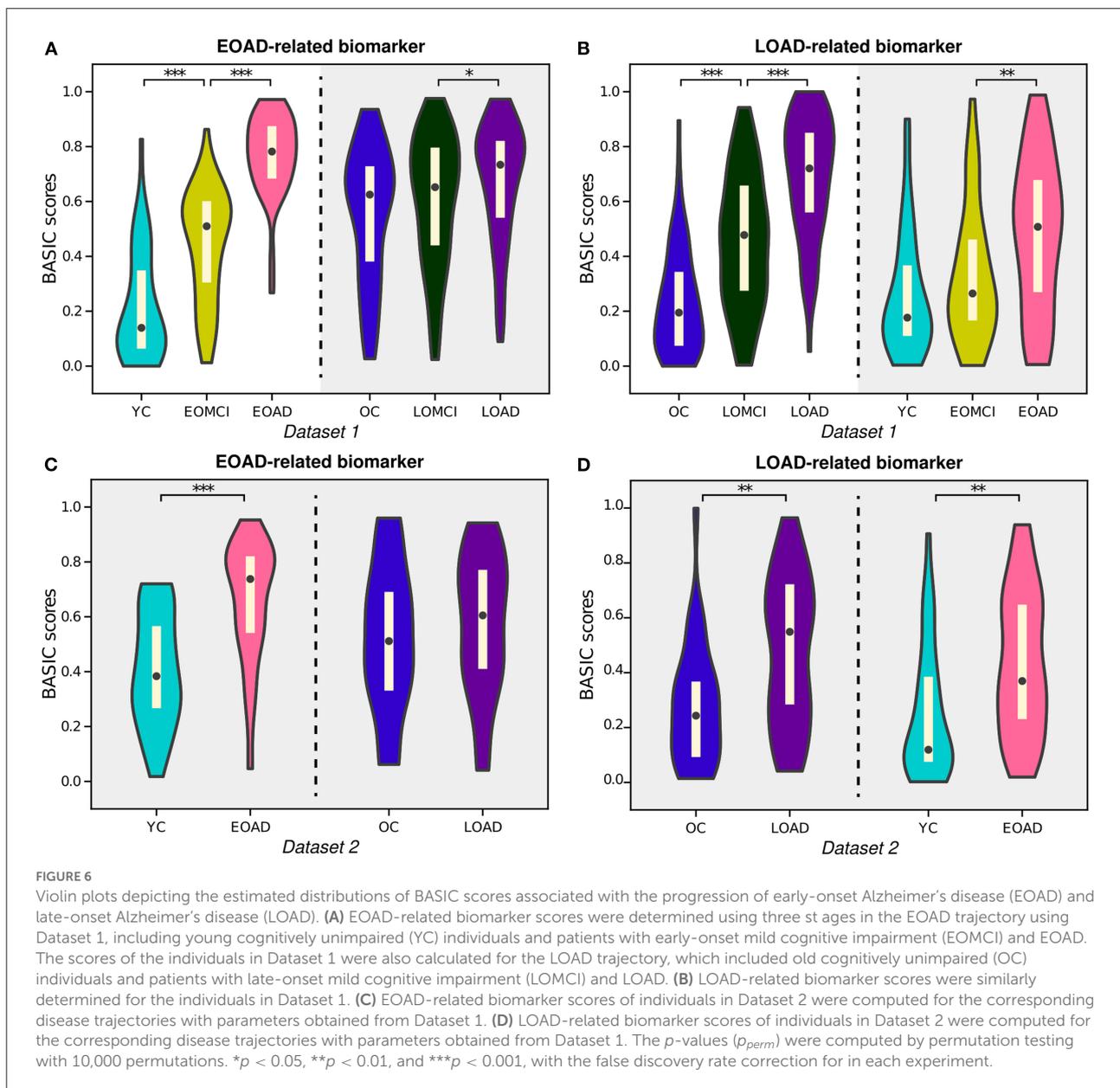
### 3.5. Construction of BASIC scores using continuous measures

There is a need to identify biomarkers associated with continuous pathological changes, which are sometimes the direct criteria for brain disorders. For instance, amyloid

deposition is a necessary condition for diagnosing AD (Jack et al., 2018); consequently, determining whether an individual is amyloid-positive or amyloid-negative is essential for distinguishing AD from other neurodegenerative disorders, which can be determined by global retention values obtained from amyloid PET. We identified a brain connectivity-based biomarker and computed its score in relation to the global amyloid SUVR (Experiment 5). The BASIC scores were significantly correlated with the SUVR (Figure 7A). In addition, the distributions of scores in the amyloid-negative ( $A\beta^-$ ) and amyloid positive ( $A\beta^+$ ) groups were smoothly connected, with scores also present near the boundary between the two groups (Figure 7B).

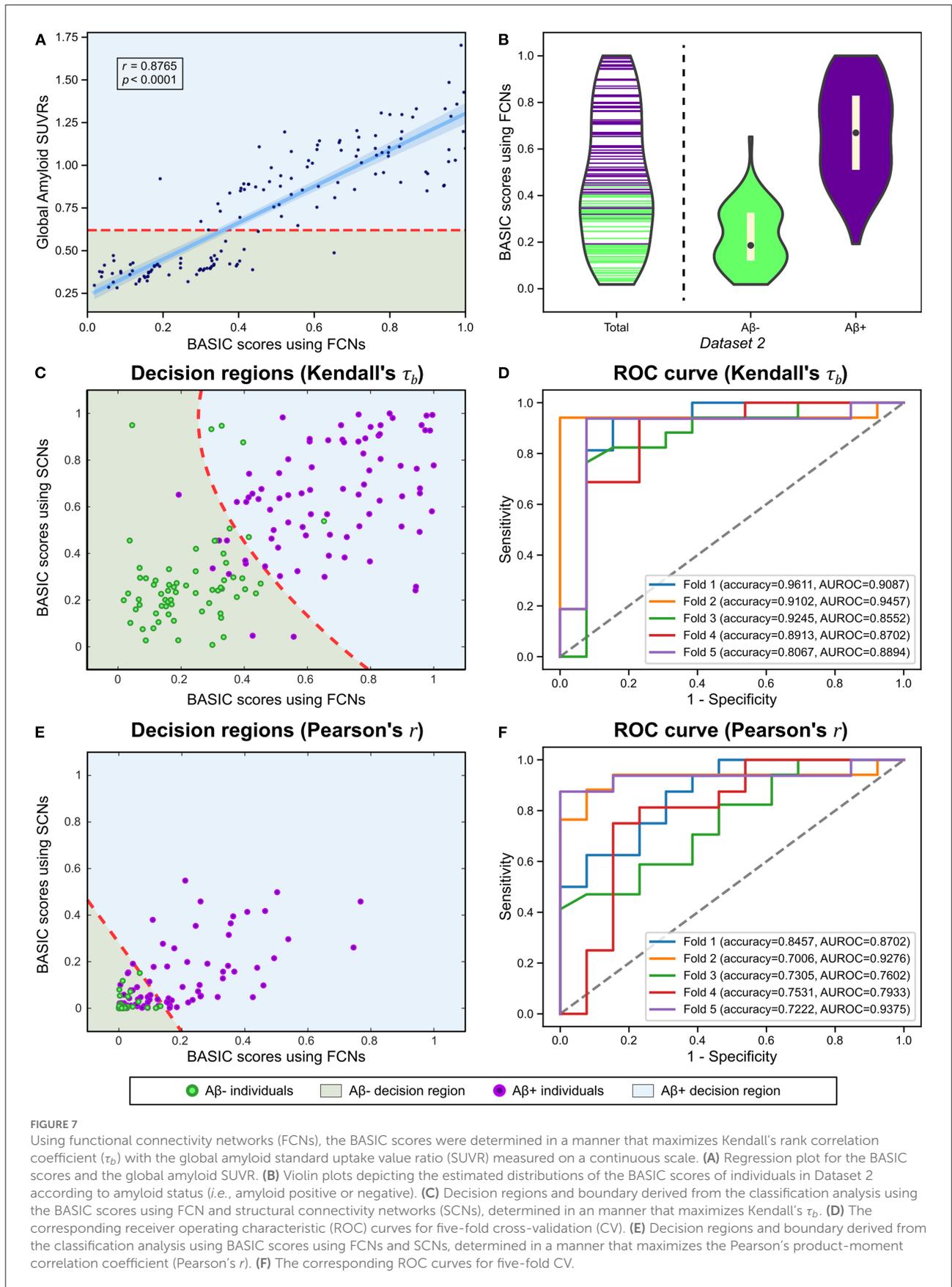
We also performed a classification analysis between  $A\beta^-$  and  $A\beta^+$  groups using BASIC scores derived from FCNs and structural connectivity networks (SCNs) from a previous study (Lee et al., 2021b). When classification was performed using BASIC scores obtained by maximizing Kendall’s  $\tau_b$ , the two groups were distinguished with a five-fold CV accuracy of 0.8987 (Figures 7C,D). However, BASIC scores obtained by maximizing the canonical correlation coefficient, or equivalently, Pearson’s  $r$  (Experiment 5-1) demonstrated poorer





disorders, summarizing these biomarkers into a single numerical value is beneficial. Previous studies have used various multivariate statistical techniques to either extract biomarkers from brain connectivity or transform them into a single score (Zalesky et al., 2010; Shen et al., 2017). We established a framework that unifies these two respective steps in a manner that maximizes the nonlinear monotonic relationship to disease stages represented either continuously or discretely, providing BASIC scores that are sensitive to disease progression as well as specific to a particular disease trajectory. In addition, the BASIC scores successfully captured continuity in brain changes along with disease progression, with superior generalizability.

It is natural that biomarker scores change continuously as the disease advances, given that the associated pathological phenomenon also changes continuously as the disease progresses. In this context, individuals with moderate conditions cannot be ignored in comparison to patients with severe conditions or healthy individuals, which should be reflected in the biomarker scores. From this perspective, the BASIC scores were not only distributed overlapping of each disease stage but were also approximately evenly distributed between the upper and lower bounds of the scores, which was not observed in the responses of the logistic models. This could be made possible by freeing the binary prediction model's assumption that the outcome follows



the Bernoulli distribution and instead maximizing the rank correlation. In addition, the BASIC score, computed by the weighted sum of connection strengths and transformation by the normal CDF, approximately followed a uniform distribution according to the central limit theorem; thus, the proportion of people with intermediate scores was greater than those with intermediate binary prediction model responses.

Our framework was designed to be generalizable in providing connectivity-based biomarker abnormality scores. This is mainly accomplished by employing a regularization procedure, in which additional prior information is introduced to obtain parameters and subsequently construct scores. First, the structural attributes inherent to brain connectivity were considered in the form of Laplacian regularization, which produces a smoothing effect on adjacent edges. If one of the adjacent edges has extreme connection strengths, its parameter will have a value similar to that of the adjacent edges. Hence, their connection strengths were affected at a similar level when constructing the score. Comparing five-fold CV performances with and without Laplacian regularization in our framework's objective function, we empirically validated that Laplacian regularization produced superior generalized performance in reflecting the stage of the disease (*Experiment 1 vs. 1'*, *Experiment 2 vs. 2'*, and *Experiment 3 vs. 3'* in Table 3). Controlling the extent of a connectivity-based biomarker and choosing the most parsimonious one based on the ISE rule also reduced the influence of unseen connectivity by decreasing the number of variables used in the score computation. In addition, the framework measured the association with the reference variable using Kendall's rank correlation, which may affect generalizability by mitigating the effect of connection strengths caused by measurement errors. Finally, the framework selects optimal hyperparameter values based on CV, which prevents data overfitting (Hastie et al., 2009; Shen et al., 2017).

Identifying reliable biomarkers, regardless of their abnormality score, is the focus of this framework. It is also possible, for instance, to estimate parameters by maximizing the association through explicit regularization while controlling the sparsity for the entire edge set. However, the collection of selected edges is inappropriate for use as a biomarker because a sparsity-controlling procedure may not produce a unique solution. The framework proposed in this study identified a biomarker with a connectivity-specific attribute by leveraging a NBS, which identifies the effect of interest with high statistical power (Zalesky et al., 2010). Moreover, given that brain connectivity should be viewed as a complex system that regulates neural information flow, the biomarker extracted by the multivariate technique combined with mass-univariate hypothesis testing can explain network-topological alterations, as in previous research (Myung et al., 2016; Kim et al., 2017, 2019). This is why our framework is

not designed as a single method but instead consists of two separable techniques.

There are several limitations to this study. We used the continuum of clinical manifestation of AD as an approximation of the AD continuum when considering continuous brain alterations for demonstrating the continuity of BASIC scores. Despite this, it is believed that demonstrating the capability of our framework is not a concern because it maximizes the association to any disease stage represented as ordinal scales. In addition, although BASIC scores capture nonlinear monotonic relationships through Kendall's  $\tau_b$ , it is limited in its ability to accurately reflect alterations in brain connectivity as the disease progresses. Suppose that multiple connected components are extracted in the identification step, showing distinct dynamic patterns of changes. In this case, linear summarization is insufficient to reflect the complex nonlinear relationship with the behavior of the biomarker. As AD progresses, the default mode network (DMN) and salient network exhibit an inverse relationship with functional connectivity (Brier et al., 2012; Zhou and Seeley, 2014). Furthermore, functional connectivity in DMN regions has been reported to have a quadratic relationship with disease progression (Schultz et al., 2017); however, this tendency cannot be measured using Kendall's  $\tau_b$ . In the future, we intend to use graph-based deep-learning models to capture and summarize these complex nonlinear relationships (Kipf and Welling, 2017; Xu et al., 2018).

In conclusion, our unified framework provides an explainable connectivity-based biomarker and continuity-reflecting BASIC scores that are sensitive and specific to disease progression. The selection of NBS statistics during the identification step was flexible within the framework. In addition, the framework is not restricted by the type of specific association measure, thresholding criterion, or summarizing function used in the summarization step. The framework is applicable to other network-based neurodegenerative disorders, such as Parkinson's disease, progressive supranuclear palsy, and amyotrophic lateral sclerosis, where brain connectivity is disrupted as the diseases progress (Brown et al., 2017; Brundin and Melki, 2017; Romano et al., 2022), although this paper focuses on AD-related brain connectivity. With this flexibility, our framework establishes a milestone for analyzing complex brain connectivity networks to explain or predict neurodegenerative disorders by presenting indicators of the degree to which brain networks collapse or by utilizing them as one of the features for predictive models.

## Data availability statement

The raw data and/or codes supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Boards (IRB) of either Samsung Medical Center or Gachon University Gil Medical Center. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

J-KS had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. J-KS and S-WK planned the study design, concept, interpreted the results, and wrote the original manuscript. HK, YN, SS, and DN collected image data. J-KS, S-WK, and Y-HS preprocessed image data and performed image-based analyses. All authors reviewed and edited the manuscript.

## Funding

This research was supported by the National Institute of Health research project (2022-ER1003-00), a grant

from the Korea Healthcare Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), and also funded by the Ministry of Health and Welfare of the Republic of Korea (Grant No H114C1135) and a National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) (No. 2018M3C7A1056889).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Alfons, A., Croux, C., and Filzmoser, P. (2017). Robust maximum association estimators. *J. Am. Stat. Assoc.* 112, 436–445. doi: 10.1080/01621459.2016.1148609
- Ando, R., and Zhang, T. (2006). "Learning on graph with laplacian regularization," in *Advances in Neural Information Processing Systems, Vol. 19* (Vancouver, BC).
- Bang, H. (2020). Biomarker score in risk prediction: beyond scientific evidence and statistical performance. *Diabetes Metab. J.* 44, 245–247. doi: 10.4093/dmj.2020.0073
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Brier, M. R., Thomas, J. B., Snyder, A. Z., Benzinger, T. L., Zhang, D., Raichle, M. E., et al. (2012). Loss of intranetwork and internetwork resting state functional connections with Alzheimer's disease progression. *J. Neurosci.* 32, 8890–8899. doi: 10.1523/JNEUROSCI.5698-11.2012
- Brossart, D. F., Laird, V. C., and Armstrong, T. W. (2018). Interpreting kendall's tau and tau-u for single-case experimental designs. *Cogent Psychol.* 5, 1518687. doi: 10.1080/23311908.2018.1518687
- Brown, J. A., Hua, A. Y., Trujillo, A., Attygalle, S., Binney, R. J., Spina, S., et al. (2017). Advancing functional dysconnectivity and atrophy in progressive supranuclear palsy. *Neuroimage Clin.* 16, 564–574. doi: 10.1016/j.nicl.2017.09.008
- Brundin, P., and Melki, R. (2017). Prying into the prion hypothesis for Parkinson's disease. *J. Neurosci.* 37, 9808–9818. doi: 10.1523/JNEUROSCI.1788-16.2017
- Cao, M., Wang, J.-H., Dai, Z.-J., Cao, X.-Y., Jiang, L.-L., Fan, F.-M., et al. (2014). Topological organization of the human brain functional connectome across the lifespan. *Dev. Cogn. Neurosci.* 7, 76–93. doi: 10.1016/j.dcn.2013.11.004
- Caravaglios, G., Muscoso, E., Blandino, V., Di Maria, G., Gangitano, M., Graziano, F., et al. (2022). EEG resting-state functional networks in amnesic mild cognitive impairment. *Clin. EEG Neurosci.* doi: 10.1177/15500594221110036. [Epub ahead of print].
- Chok, N. S. (2010). *Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data* (Ph.D. thesis). University of Pittsburgh.
- Couso, I., Strauss, O., and Saulnier, H. (2018). Kendall's rank correlation on quantized data: an interval-valued approach. *Fuzzy Sets Syst.* 343, 50–64. doi: 10.1016/j.fss.2017.09.003
- croux, C., Filzmoser, P., and Oliveira, M. R. (2007). Algorithms for projection-pursuit robust principal component analysis. *Chemometr. Intell. Lab. Syst.* 87, 218–225. doi: 10.1016/j.chemolab.2007.01.004
- Fox, M. D., Zhang, D., Snyder, A. Z., and Raichle, M. E. (2009). The global signal and observed anticorrelated resting state brain networks. *J. Neurophysiol.* 101, 3270–3283. doi: 10.1152/jn.90777.2008
- Freedman, D., and Lane, D. (1983). A nonstochastic interpretation of reported significance levels. *J. Bus. Econ. Stat.* 1, 292–298. doi: 10.1080/07350015.1983.10509354
- Gomez-Ramirez, J., and Wu, J. (2014). Network-based biomarkers in Alzheimer's disease: review and future directions. *Front. Aging Neurosci.* 6, 12. doi: 10.3389/fnagi.2014.00012
- Grady, C. L., Furey, M. L., Pietrini, P., Horwitz, B., and Rapoport, S. I. (2001). Altered brain functional connectivity and impaired short-term memory in Alzheimer's disease. *Brain* 124, 739–756. doi: 10.1093/brain/124.4.739
- Grömping, U. (2015). Variable importance in regression models. *Wiley Interdisc. Rev. Comput. Stat.* 7, 137–152. doi: 10.1002/wics.1346
- Han, C. E., Yoo, S. W., Seo, S. W., Na, D. L., and Seong, J.-K. (2013). Cluster-based statistics for brain connectivity in correlation with behavioral measures. *PLoS ONE* 8, e72332. doi: 10.1371/journal.pone.0072332

- Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology* 143, 29–36. doi: 10.1148/radiology.143.1.7063747
- Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Vol. 2*. New York, NY: Springer.
- Jack, C. R. Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., et al. (2018). NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* 14, 535–562. doi: 10.1016/j.jalz.2018.02.018
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015
- Joubert, S., Gour, N., Guedj, E., Didic, M., Guériot, C., Koric, L., et al. (2016). Early-onset and late-onset Alzheimer's disease are associated with distinct patterns of memory impairment. *Cortex* 74, 217–232. doi: 10.1016/j.cortex.2015.10.014
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika* 30, 81–93. doi: 10.1093/biomet/30.1-2.81
- Kim, H. J., Cha, J., Lee, J.-M., Shin, J. S., Jung, N.-Y., Kim, Y. J., et al. (2016). Distinctive resting state network disruptions among Alzheimer's disease, subcortical vascular dementia, and mixed dementia patients. *J. Alzheimers Dis.* 50, 709–718. doi: 10.3233/JAD-150637
- Kim, J. E., Kim, S.-W., Choi, M., Seong, J.-K., and Lee, J.-H. (2019). Identifying brain connectivity using network-based statistics in amnesic mild cognitive impairment stratified by  $\beta$ -amyloid positivity. *Am. J. Alzheimers Dis. Other Dementias*, *textregistered* 34, 104–111. doi: 10.1177/1533317518813556
- Kim, K., Kim, S.-W., Myung, W., Han, C. E., Fava, M., Mischoulon, D., et al. (2017). Reduced orbitofrontal-thalamic functional connectivity related to suicidal ideation in patients with major depressive disorder. *Sci. Rep.* 7, 1–11. doi: 10.1038/s41598-017-15926-0
- Kim, Y., Kim, S.-W., Seo, S. W., Jang, H., Kim, K. W., Cho, S. H., et al. (2021). Effect of education on functional network edge efficiency in Alzheimer's disease. *Sci. Rep.* 11, 1–8. doi: 10.1038/s41598-021-96361-0
- Kipf, T. N., and Welling, M. (2017). “Semi-supervised classification with graph convolutional networks,” in *International Conference on Learning Representations (ICLR)* (Toulon).
- Lee, J.-J., Kim, H. J., Ćeko, M., Park, B.-Y., Lee, S. A., Park, H., et al. (2021a). A neuroimaging biomarker for sustained experimental and clinical pain. *Nat. Med.* 27, 174–182. doi: 10.1038/s41591-020-1142-7
- Lee, J. S., Kim, C., Shin, J.-H., Cho, H., Shin, D.-S., Kim, N., et al. (2018). Machine learning-based individual assessment of cortical atrophy pattern in Alzheimer's disease spectrum: development of the classifier and longitudinal evaluation. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-018-22277-x
- Lee, W. J., Yoon, C. W., Kim, S.-W., Jeong, H. J., Seo, S., Na, D. L., et al. (2021b). Effects of Alzheimer's and vascular pathologies on structural connectivity in early- and late-onset Alzheimer's disease. *Front. Neurosci.* 15, 58. doi: 10.3389/fnins.2021.606600
- Li, M., Zheng, G., Zheng, Y., Xiong, Z., Xia, R., Zhou, W., et al. (2017). Alterations in resting-state functional connectivity of the default mode network in amnesic mild cognitive impairment: an fMRI study. *BMC Med. Imaging* 17, 1–8. doi: 10.1186/s12880-017-0221-9
- Lindquist, M. A. (2008). The statistical analysis of fMRI data. *Stat. Sci.* 23, 130–157. doi: 10.1214/09-STS282
- Mendez, M. F. (2019). Early-onset Alzheimer disease and its variants. *Continuum* 25, 34. doi: 10.1212/CON.0000000000000687
- Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B., and Bandettini, P. A. (2009). The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? *Neuroimage* 44, 893–905. doi: 10.1016/j.neuroimage.2008.09.036
- Myung, W., Han, C., Fava, M., Mischoulon, D., Papakostas, G., Heo, J., et al. (2016). Reduced frontal-subcortical white matter connectivity in association with suicidal ideation in major depressive disorder. *Transl. Psychiatry* 6, e835–e835. doi: 10.1038/tp.2016.110
- Panegyres, P. K., and Chen, H.-Y. (2013). Differences between early and late onset Alzheimer's disease. *Am. J. Neurodegener. Dis.* 2:300. doi: 10.3389/fneur.2012.00081
- Poline, J.-B., Worsley, K. J., Evans, A. C., and Friston, K. J. (1997). Combining spatial extent and peak intensity to test for activations in functional imaging. *Neuroimage* 5, 83–96. doi: 10.1006/nimg.1996.0248
- Romano, A., Lopez, E. T., Liparoti, M., Polverino, A., Minino, R., Trojsi, F., et al. (2022). The progressive loss of brain network fingerprints in amyotrophic lateral sclerosis predicts clinical impairment. *Neuroimage Clin.* 35, 103095. doi: 10.1016/j.nicl.2022.103095
- Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069. doi: 10.1016/j.neuroimage.2009.10.003
- Schultz, A. P., Chhatwal, J. P., Hedden, T., Mormino, E. C., Hanseeuw, B. J., Sepulcre, J., et al. (2017). Phases of hyperconnectivity and hypoconnectivity in the default mode and salience networks track with amyloid and tau in clinically normal individuals. *J. Neurosci.* 37, 4323–4331. doi: 10.1523/JNEUROSCI.3263-16.2017
- Schumacher, J., Gunter, J. L., Przybelski, S. A., Jones, D. T., Graff-Radford, J., Savica, R., et al. (2021). Dementia with Lewy bodies: association of Alzheimer pathology with functional connectivity networks. *Brain* 144, 3212–3225. doi: 10.1093/brain/awab218
- Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., et al. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat. Protoc.* 12, 506–518. doi: 10.1038/nprot.2016.178
- Shmueli, G. (2010). To explain or to predict? *Stat. Sci.* 25, 289–310. doi: 10.1214/10-STS330
- Sun, L., Ji, S., and Ye, J. (2008). “A least squares formulation for canonical correlation analysis,” in *Proceedings of the 25th International Conference on Machine Learning (Helsinki)*, 1024–1031.
- Tellechea, P., Pujol, N., Esteve-Belloch, P., Echeveste, B., García-Eulate, M., Arbizu, J., et al. (2018). Early- and late-onset Alzheimer disease: are they the same entity? *Neurologia (English Edition)* 33, 244–253. doi: 10.1016/j.nrleng.2015.08.009
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Wang, K., Liang, M., Wang, L., Tian, L., Zhang, X., Li, K., et al. (2007). Altered functional connectivity in early Alzheimer's disease: a resting-state fMRI study. *Hum. Brain Mapp.* 28, 967–978. doi: 10.1002/hbm.20324
- Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage* 92, 381–397. doi: 10.1016/j.neuroimage.2014.01.060
- Worsley, K. J. (2001). Testing for signals with unknown location and scale in a  $\chi^2$  random field, with an application to fMRI. *Adv. Appl. Probab.* 33, 773–793. doi: 10.1239/aap/1011994029
- Xu, K., Hu, W., Leskovec, J., and Jegelka, S. (2018). “How powerful are graph neural networks?” in *International Conference on Learning Representations (ICLR)* (Vancouver, BC).
- Zalesky, A., Cocchi, L., Fornito, A., Murray, M. M., and Bullmore, E. (2012). Connectivity differences in brain networks. *Neuroimage* 60, 1055–1062. doi: 10.1016/j.neuroimage.2012.01.068
- Zalesky, A., Fornito, A., and Bullmore, E. T. (2010). Network-based statistic: identifying differences in brain networks. *Neuroimage* 53, 1197–1207. doi: 10.1016/j.neuroimage.2010.06.041
- Zhan, Y., Yao, H., Wang, P., Zhou, B., Zhang, Z., An, N., et al. (2016). Network-based statistic show aberrant functional connectivity in Alzheimer's disease. *IEEE J. Sel. Top. Signal Process.* 10, 1182–1188. doi: 10.1109/JSTSP.2016.2600298
- Zhou, B., Liu, Y., Zhang, Z., An, N., Yao, H., Wang, P., et al. (2013). Impaired functional connectivity of the thalamus in Alzheimer's disease and mild cognitive impairment: a resting-state fMRI study. *Curr. Alzheimer Res.* 10, 754–766. doi: 10.2174/15672050113109990146
- Zhou, J., and Seeley, W. W. (2014). Network dysfunction in Alzheimer's disease and frontotemporal dementia: implications for psychiatry. *Biol. Psychiatry* 75, 565–573. doi: 10.1016/j.biopsych.2014.01.020