ORIGINAL ARTICLE

SMART (Splitting-Merging Assisted Reliable) Independent Component Analysis for Extracting Accurate Brain Functional Networks

Xingyu He[1] · Vince D. Calhoun[2] · Yuhui Du[1,2]

© Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences 2024

Abstract Functional networks (FNs) hold significant promise in understanding brain function. Independent component analysis (ICA) has been applied in estimating FNs from functional magnetic resonance imaging (fMRI). However, determining an optimal model order for ICA remains challenging, leading to criticism about the reliability of FN estimation. Here, we propose a SMART (splitting-merging assisted reliable) ICA method that automatically extracts reliable FNs by clustering independent components (ICs) obtained from multi-model-order ICA using a simplified graph while providing linkages among FNs deduced from different-model orders. We extend SMART ICA to multi-subject fMRI analysis, validating its effectiveness using simulated and real fMRI data. Based on simulated data, the method accurately estimates both group-common and group-unique components and demonstrates robustness to parameters. Using two age-matched cohorts of resting fMRI data comprising 1,950 healthy subjects, the resulting reliable group-level FNs are greatly similar between the two cohorts, and interestingly the subject-specific FNs show progressive changes while age increases. Furthermore, both small-scale and large-scale brain FN templates are provided as benchmarks for future studies. Taken together, SMART ICA can automatically obtain reliable FNs in analyzing multi-subject fMRI data, while also providing linkages between different FNs.

Introduction

Functional networks (FNs) derived from functional magnetic resonance imaging (fMRI) data represent integrative and interactive relations among different brain regions [1] and are promising in disclosing information about human brain function [2]. Many studies have explored brain FNs in healthy populations [3–5] and in the mechanisms of mental disorders [6–9] since FNs play a pivotal role in elucidating cognitive activities in the human brain and identifying reliable biomarkers at the network level. Therefore, developing reliable methods of brain FN analysis is particularly critical for maximizing the robustness and generalizability of the findings on brain FNs.

Although different approaches have been used to extract brain FNs [10, 11], data-driven independent component analysis (ICA) is one of the most widely applied methods for understanding healthy brain function [12, 13] as well as exploring brain disorders [14–17]. Relative to hypothesis-based methods, such as seed-based or region-of-interest-based methods, ICA does not require a priori determination of seeds or brain regions. Spatial ICA decomposes fMRI data into spatially independent components, some of which represent meaningful FNs while others could be artifacts. Using spatial ICA, different components are modeled to have no or less spatial overlap, measured by independence,

✉ Yuhui Du
duyuhui@sxu.edu.cn

1 School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

2 Tri-Institutional Center for Translational Research in Neuroimaging and Data Science, Georgia State University, Georgia Institute of Technology, Emory University, Atlanta 30303, USA

to obtain FNs that handle different functions. In practice, ICA also can help explore intricate functional associations as a certain brain region may be simultaneously involved in the operation of multiple brain functions [18, 19]. Although ICA has been successful in extracting networks, it has some inherent shortcomings in extracting FNs, especially in terms of the determination of model order (i.e., the number of components). The uncertainty of the model order hinders the capability of ICA in FN extraction and subsequent exploration of biomarkers to some extent.

Indeed, some early studies have suggested estimating the number of components in ICA using information theoretic criteria (ITC) [20], such as the minimum description-length, Akaike's information, and Bayesian information criteria. Based on the finite memory length and the autoregressive model, entropy-rate-based model order selection methods using ITC have been proposed to utilize all available subjects' data [21]. Some studies have also used information theory methods to estimate the model order of ICA in fMRI applications [22–24]. However, the estimation accuracy of these methods is compromised by the presence of complex noise structures [25]. Despite various methods being proposed, their results may vary significantly, and no method can guarantee an accurate estimate of the number of components.

A few studies avoid setting a specific model order of ICA but estimate relatively robust FNs by changing the parameter settings and the data input. Kuang *et al.* estimated FNs with different model orders and selected the 'best' result that fit well with the reference networks [26]. However, the method relies heavily on the selection of reference networks. A method called Snowball ICA first generates seed components by applying ICA to randomly selected subjects' fMRI data and then updates the seed components iteratively by adding different blocks of fMRI data until all subjects' data are used [27]. Although the method does not require the input of the component number, it is sensitive to the generation of the seed components and the organized arrangement of fMRI blocks. Furthermore, previous methods only focus on extracting one group of FNs, without providing any information to depict the relationship among FNs of different scales. Another study suggests that FN connectivity between FNs obtained from multi-model-order ICA can provide more intriguing findings than single-model-order ICA [28]; however, it only aggregates all FN connectivity strengths from different orders and does not truly combine them.

In this paper, we propose a method, named SMART (splitting-merging assisted reliable) ICA, which achieves the automatic estimation of reliable FNs without the need to require a specific model order and meanwhile provides the linkage relationship among networks with different scales. The main contributions are as follows: (1) Our method effectively combines a clustering technique with

ICA to automatically extract reliable FNs from multi-model-order ICA results, which facilitates the use of ICA in brain FN analysis since determining an optimal model order is often difficult for traditional ICA. (2) We develop a splitting-merging clustering approach that not only iteratively identifies the optimal cluster centers during the splitting-merging process but also provides the linkage information among the independent components (ICs) obtained from different model orders by making full use of the FNs constructed in a tree structure. Different from previous studies, our method allows for a comprehensive understanding of the FNs under various model orders. (3) We propose a scheme to extend SMART ICA to multiple-subject analysis and validate our method using both simulated and real fMRI data. Using multiple simulated datasets with different properties, our method outputs accurate subject-specific FNs and is insensitive to the parameter setting. When applied to real fMRI data, SMART ICA yields highly consistent FNs between two age-matched healthy cohorts. Interestingly, our method detects subtle and progressive functional changes in the brains of a healthy population associated with increasing age. (4) Based on the large-sample fMRI data from 1,950 healthy subjects, we construct reliable FN templates at both the small- and large-scale, which provide an important benchmark for future FN studies using fMRI data.

Materials and Methods

Here, we provide a detailed description of the SMART ICA method and its evaluation process. We propose a splitting- and merging-assisted clustering algorithm in conjunction with a graph simplification technique to automatically cluster ICs from multi-model-order ICA runs on fMRI data. SMART ICA not only yields reliable and meaningful FNs represented by cluster centers but also establishes linkage relationships between different-scale ICs resulting from multi-model-order ICA. It is worth pointing out that SMART ICA applies to both individual-subject and multi-subject data analyses. It is known that ICA on individual-subject data may result in ICs with a random order [29], and a common solution is to obtain reliable group-level ICs and then estimate subject-specific ICs based on them to facilitate the analysis of multi-subject data in fMRI studies. So, in the following, we primarily describe the extension of SMART ICA to the multi-subject fMRI data analysis. All experiments in this paper were performed on MATLAB 2018a and 2022a.

Our SMART (Splitting-Merging-Assisted Reliable) ICA Method

As shown in Fig. 1A, multi-subject fMRI data analysis using SMART ICA primarily includes four steps. First, initial group-level ICs are obtained by applying ICA [30] to the entire multi-subject fMRI data using different model orders. Second, we apply the proposed splitting- and merging-assisted clustering technique to cluster the initial group-level ICs, resulting in reliable group-level ICs. Simultaneously, this process establishes linkage relationships between the initial group-level ICs. Third, artifact-related ICs are removed from reliable group-level ICs, retaining reliable group-level FNs. Finally, subject-specific FNs are computed by applying group information-guided ICA (GIG-ICA) [14, 31] with the guidance of reliable group-level FNs.

(1) ICA with Different Model Orders

In step 1, multi-model-order ICA runs on the fMRI data of multiple subjects (e.g., P subjects) are performed to obtain the initial group-level ICs. Using each model order (i.e., a specific component number N), we conduct the following procedure for obtaining group-level ICs. Firstly, we transform the pth subject's fMRI data into a matrix $X_p \in \mathbb{R}^{T \times M}$, where T represents the number of time points and M represents the number of voxels within a common brain mask. Based on the data of each subject ($X_p, p = 1, 2, \ldots, P$), the subject-level principal component analysis (PCA) is conducted for dimensionality reduction along the time point direction. Then, the reduced data of all subjects are concatenated along the time point direction and the group-level PCA is carried out for a further dimensionality reduction, resulting in a matrix $H \in \mathbb{R}^{N \times M}$. Here, N is the number of components. Finally, the ICA (Infomax algorithm in this paper) [32, 33] with an additional stabilization technique [34] is implemented to decompose H into N ICs. ICA is formulated by Eq. (1).

$$H \approx Q \times S, \tag{1}$$

here, $Q \in \mathbb{R}^{N \times N}$ denotes the mixing matrix, and $S \in \mathbb{R}^{N \times M}$ denotes the group-level ICs. Our method sets N to different numbers to represent different model orders, and a total of g initial group-level ICs are obtained across different model orders.

(2) Splitting- and Merging-Assisted Clustering

In step 2, we introduce a novel splitting- and merging-assisted clustering algorithm combined with a graph simplification method to group all g initial group-level ICs into distinct clusters. All cluster centers are used to represent reliable group-level ICs. As illustrated in Fig. 1B, a graph with all initial group-level ICs as nodes is constructed and then simplified to a tree based on a previous method [35]. A forest is initialized with this tree, and an iterative splitting and merging process is done until the structures of all trees in the forest stabilize. After that, each tree in the forest is regarded as a cluster whose center is regarded as one reliable group-level IC, and the linkage relationships between initial group-level ICs within each tree are captured. The details of the clustering method are stated as follows.

Here, we employ a simplified tree-based community detection (STCD) method [35] to transform the intricate relationships among the initial group-level ICs into a simplified tree structure, serving as the foundation for our clustering process. We begin with g initial group-level ICs and create graph G to depict their relationships. Regarding $G = \langle V, E, A \rangle$, the node set V encompasses all initial group-level ICs, E signifies the set of edges between the ICs, and A represents the adjacent matrix of G. Hereafter, $\langle v_i, v_j \rangle \in E$ represents the edge between IC v_i and IC v_j, and A_{ij} that denotes the (i, j)th element of A contains the absolute value of the Pearson correlation coefficient between IC v_i and IC v_j. Next, we mine the intricate relationships among all initial group-level ICs to simplify the graph G by the STCD method. For each node (i.e., each IC), we utilize the edges between it and other nodes (i.e., other ICs) to represent the properties of the node, which are then used to quantify

Fig. 1 The pipeline of SMART ICA. **A** The pipeline of our method, which primarily includes four steps. **B** Step 2 in more detail. Here, the simplified tree-based community detection (STCD) algorithm can simplify a complex graph structure into a tree-like structure. ICA, independent component analysis; fMRI, functional magnetic resonance imaging; ICs, independent components; FNs, functional networks.

908

Neurosci. Bull. July, 2024, 40(7):905–920
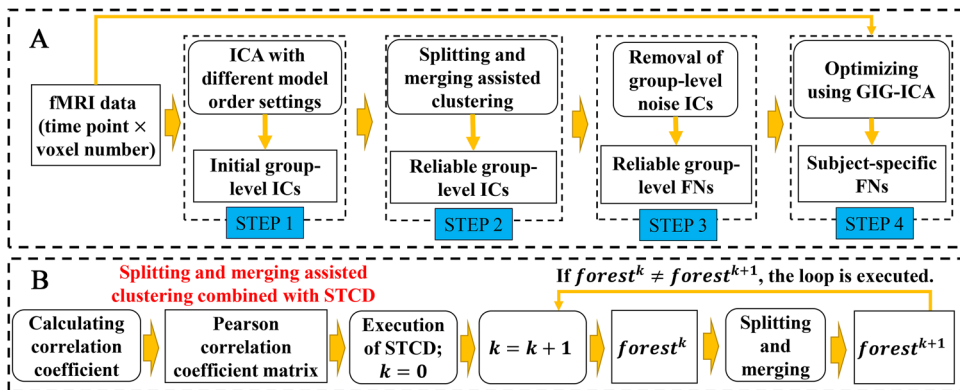
the relationships between any two ICs. As such, the similarity $\delta(v_i, v_j)$ between IC v_i and IC v_j is then formalized as:

$$\delta(v_i, v_j) = \sum_{v_z \in V, v_z \neq v_i, v_z \neq v_j} A_{iz} A_{jz}. \tag{2}$$

This calculates the similarity between two ICs on a global scale by taking into account their relations with other ICs, rather than solely focusing on the similarity between the two ICs themselves. After that, we evaluate how each IC leads or follows other ICs to further extract their important relations. For the IC v_i, its leading degree is computed by:

$$L(v_i) = \sum_{D(v_j) < D(v_i), v_j \in V, v_j \neq v_i} \delta(v_i, v_j). \tag{3}$$

Here, the degree of IC v_i [i.e., $D(v_i)$] is defined as the sum of the absolute values of the Pearson correlation coefficients between the IC v_i and all other ICs, and is formulated by:

$$D(v_i) = \sum_{v_j \in V, v_j \neq v_i} A_{ij}. \tag{4}$$

The following degree of IC v_i over IC v_j is defined as:

$$F(v_i, v_j) = \begin{cases} \frac{\delta(v_i, v_j)}{D(v_i)}, & if\ L(v_j) \geq L(v_i) \\ 0, & otherwise \end{cases}. \tag{5}$$

Based on the leading degree and following degree of each IC, a new graph with a tree structure $G' = \langle V', E', A' \rangle$ is established by:

$$\begin{cases} V' = V, \\ E' = \left\{ \langle v_i, v_q \rangle \middle| \begin{array}{c} F(v_i, v_q) = \max_{v_j \in V} F(v_i, v_j), \\ v_i \in V \end{array} \right\}, \\ A' = \left\{ A'_{ij} \middle| \begin{array}{c} A'_{ij} = F(v_i, v_j), if \langle v_i, v_j \rangle \in E' \\ A'_{ij} = 0, otherwise \end{array} \right\} \end{cases}. \tag{6}$$

A forest, which is composed of all initial group-level ICs and their relationships, is initialized using the tree structure G'. Inspired by a previous study [36], we propose the following splitting and merging method to iteratively segment the forest into multiple stable trees. Each tree within the forest corresponds to a cluster and the cluster center (C_i) of the ith cluster is defined as the IC with the maximum degree sum according to the tree structure as formulated in Eq. (7).

$$C_i = \left\{ v_x \middle| \begin{array}{c} \sum_{v_y \in V^i} A'_{yx} = max \left\{ \begin{array}{c} \sum_{v_y \in V^i} A'_{yj}, \\ j = \{j | v_j \in V^i\} \end{array} \right\}, \\ v_x \in V^i \end{array} \right\}, \tag{7}$$

where $A'y_x$ represents the (y, x) element of the A' and V^i represents the set of all ICs in ith tree.

Specifically, for the ith tree, a splitting operation is conducted if its intra-cluster distance d^i_{intra} exceeds the half of current average inter-cluster distance d_{mean} (i.e., $d^i_{intra} > d_{mean}/2$). The average inter-cluster distance and the intra-cluster distance are defined by Eqs (8) and (9), respectively.

$$d_{mean} = \begin{cases} \frac{2}{o \times (o-1)} \sum_{i=1}^{o} \sum_{j=i+1}^{o} d(C_i, C_j), & if\ o > 1 \\ 0, & if\ o = 1 \end{cases}, \tag{8}$$

where o represents the current cluster numbers and $d(C_i, C_j)$ represents the distance between C_i and C_j.

$$d^i_{intra} = \max_{v_x \in V^i} \{d(C_i, v_x)\} + \min_{v_y \neq C_i, v_y \in V^i} \{d(C_i, v_y)\}, \tag{9}$$

where the intra-cluster distance (d^i_{intra}) of the ith tree is regarded as the sum of the distance between the cluster center C_i and the nearest IC (except for itself) as well as the distance between the cluster center and the farthest IC. Here, the distance between any two ICs (such as IC v_x and IC v_y) is defined as:

$$d(v_x, v_y) = 1 - |corr(v_x, v_y)|, \tag{10}$$

where corr(·) represents the Pearson correlation coefficient, and |·| represents absolute value operation.

As for the splitting process, our method searches through the forest to identify trees that satisfy the splitting condition. Subsequently, it divides each of these trees into two separate trees by eliminating the edge with the minimum following degree. This process continues until all trees in the forest no longer meet the splitting condition.

Following the completion of the splitting operation, we proceed with the merge operation. The merging condition of the ith and jth trees is that the distance between C_i and C_j is less than half of the average inter-cluster distance [i.e., $d(C_i, C_j) < d_{mean}/2$]. we identify candidate pairs of trees that meet the merging condition. Subsequently, our method searches all candidate pairs to merge the paired trees with the minimum $d(C_i, C_j)$ into a new tree using the STCD method.

The splitting and merging operations described above are conducted iteratively until the structures of all trees reach a stable state. Consequently, we assign ICs within the same tree using the same label. The linkage relationships between the initial group-level ICs included in each tree are retained and each cluster center is regarded as one reliable group-level IC.

(3) Removal of Group-level Artifact ICs

In step 3, based on the reliable group-level ICs yielded from step 2, the artifact ICs are removed and the remaining reliable group-level ICs are taken as reliable group-level FNs. For the simulated data, we propose a method to detect artifact ICs by measuring the smoothness of reliable group-level ICs reflected by the number of Maximally Stable Extremal Regions (MSERs) [37–39]. Here, we retain those ICs with < 150 MSERs, while removing ICs with a higher count. For the real fMRI data, due to the small number of reliable group-level ICs, we manually remove the artifact ICs, such as the ICs with peak activation located in white matter and cerebrospinal fluid [40, 41].

(4) Estimation of Subject-specific Functional Networks Using the Group Information-Guided ICA (GIG-ICA) Method

In step 4, GIG-ICA [31] with reliable group-level FNs as the guide information is performed based on the fMRI data of each subject, which results in subject-specific FNs and related time courses (TCs). While various methods, such as PCA-based back-reconstruction [30] and dual regression [42], can be used to estimate subject-specific FNs, GIG-ICA [43] stands out for its superior performance due to its optimization of the independence of subject-specific FNs. Consequently, we employ GIG-ICA as the method of choice for estimating individual FNs and corresponding TCs here.

Validation Using Simulated Data

In this section, we evaluate SMART ICA using the simulated data. Two groups of simulated data with both common and unique real spatial mapping (SMs) to evaluate if our SMART ICA method can extract accurate subject-specific FNs from data with different properties. Six groups of data with different numbers of real SMs were simulated to evaluate the sensitivity of SMART ICA for the parameter settings.

(1) Evaluation of SMART ICA Based on Simulated Data with Both Common and Unique SMs

To test the utility of the SMART ICA method, we designed experiments using two groups of simulated data with both group-common and group-unique SMs, applying SMART ICA and assessing its performance. Three main aspects were evaluated: the clustering capability for the initial group-level ICs, the ability to capture linkage information between different-scale ICs, and the similarity between the extracted subject-specific FNs and the real SMs.

Two groups (Group 1 and Group 2) of simulated data with both common and unique SMs were generated *via* the SimTB toolbox [44]. Each group included 100 subjects, while each subject's data were generated using 8 SMs with small spatial overlaps and related TCs (with 300 time points for each TC). Among the 8 SMs, a common template was used for each of the six SMs, and different templates were used for each of the other two SMs to simulate the variability between the two groups. Each SM had 148×148 voxels. To simulate the subject variation, the *x*-transition and *y*-transition with the mean value = 0 voxel and SD = 1 voxel, the rotation with the mean value = 0° and SD = 1°, as well as the spread with the mean value = 1 and the SD = 0.01 were added for each SM. Finally, additional noise (with the signal-to-noise ratio = 1) was also added to the two groups of simulated data.

Based on the generated simulated data with both common and unique SMs, we conducted the following process. In step 1, for each given model order N, ICA was applied to all data of the two groups to obtain initial group-level ICs. In the experiment, N was set from 4 to 14 with a step of 2, so g (54) initial group-level ICs were obtained. For step 2, our clustering method clustered g initial group-level ICs to obtain reliable group-level ICs. In step 3, artifact ICs were removed from the reliable group-level ICs, resulting in reliable group-level FNs. In step 4, based on the remaining reliable group-level FNs and the simulated data of each subject, GIG-ICA was applied to compute subject-specific FNs and corresponding TCs.

After that, we assessed the performance of the SMART ICA method from three perspectives. We initially assessed whether the initial group-level ICs were effectively grouped using the proposed clustering method. Effective clustering tends to exhibit greater intra-cluster compactness and inter-cluster separability. In our work, the absolute values of Pearson correlation coefficients between initial group-level ICs that were sorted according to the cluster labels were calculated to demonstrate the intra-cluster and inter-cluster similarity. We also calculated the mean of the absolute values of the Pearson correlation coefficients between paired initial group-level ICs for each tree to measure intra-cluster compactness. Moreover, we calculated the absolute values of Pearson correlation coefficients among reliable group-level FNs to explore whether they are unique.

SMART ICA had an advantage in capturing linkage relationships (i.e., following and being followed) between different ICs. Therefore, we visualized the linkages between ICs within each cluster to explore the relationships between ICs obtained under different model orders. It was expected that each cluster would contain similar ICs obtained under different model orders.

Since data from two groups were simulated using both group-common and group-specific SMs but were analyzed as a whole, we evaluated the similarities between subject-specific FNs/TCs and real SMs/TCs to verify our method in

capturing subject variability. We first matched the subject-specific FNs and ground-truth SMs of all subjects using a greedy spatial correlation analysis [14] according to the absolute value of the Pearson correlation coefficient between them. Then, the Pearson correlation coefficients between the matched FNs/TCs and SMs/TCs were used to represent similarities. Finally, the similarities of FNs/TCs across all subjects were displayed using boxplots.

(2) Evaluation of SMART ICA Based on Simulated Data with Different Numbers of SMs

To further validate the sensitivity of SMART ICA for the parameter settings, we designed experiments using six groups of simulated data with varying numbers of SMs, applying SMART ICA and assessing its performance. Six groups of simulated data with varying numbers of SMs were generated *via* the SimTB toolbox [44] and the number of real SMs in the six groups was 4, 6, 8, 10, 12, and 14, respectively. Each group included 50 subjects and other data generation parameters were the same as those in the above section (1).

Based on the simulated data generated with varying numbers of SMs, we conducted independent experiments using the same parameter setting. For each of the six groups of simulated data, we separately applied the SMART ICA method to obtain initial group-level ICs, reliable group-level ICs, reliable group-level FNs, and subject-specific FNs/TCs. It is worth pointing out that the same model order set including 4, 6, 8, 10, 12, and 14 was used for the six groups of data while performing the multi-model-order ICA runs in the SMART ICA. We were interested in investigating whether SMART ICA can work well under different conditions.

To assess the sensitivity of the method to parameter settings, we primarily focused on the number of reliable group-level FNs, as well as the similarities between subject-specific FNs/TCs and real SMs/TCs. The closer the number of reliable group-level FNs is to the number of real SMs, and the higher the similarity between subject-specific FNs/TCs and real SMs/TCs, the less sensitive SMART ICA is to the parameter settings. For reliable group-level FNs of each group, we calculated the numbers of reliable group-level ICs and FNs and summarized them. For subject-specific FNs/TCs of each group, we matched them with the true SMs/TCs using a greedy spatial correlation analysis [14] according to the absolute value of the Pearson correlation coefficient between them, obtaining the similarities between extracted FNs/TCs and real SMs/TCs for each subject. We averaged the similarities of all FNs/TCs within a subject to represent the FN/TC similarity of the subject. Here, we used boxplots to demonstrate the FN/TC similarity for all subjects across all six groups.

Validation Using fMRI Data

Here, we evaluated SMART ICA using real fMRI data of large-sample healthy cohorts. Two groups of fMRI data collected from healthy populations with similar demographic characteristics were used to test the reproducibility of the results using our method. Importantly, the low-model-order and high-model-order ranges were set separately to perform SMART ICA, aiming to provide both small- and large-scale FNs for validation. We assessed the clustering performance of our method on the initial group-level ICs and the ability of our method to capture linkage relationships between different-scale ICs. In addition, we verified the correspondence and specificity of estimated FNs across all subjects, with an interest in investigating if our method can identify subtle changes in FNs along with increasing age.

(1) Materials

We analyzed the fMRI data of the subjects aged from 45 to 55 in the UK BioBank project [45]. UK Biobank data has approval from the North West Multi-Centre Research Ethics Committee as a Research Tissue Bank (please see https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics for details). This research has been conducted with the UK Biobank Resource under the project: Application ID: 34175, Applicant PI: Yuhui Du. For each subject's data, we removed the first 10 time points and then applied the rigid body motion correction to correct the subject's head motion, followed by the slice-timing correction to account for the timing difference in slice acquisition. fMRI data were subsequently warped into the standard Montreal Neurological Institute space using an echo planar imaging template and were then resampled to 3 mm $\times$ 3 mm $\times$ 3 mm isotropic voxels. The resampled fMRI images were further smoothed using a Gaussian kernel with a full width at half maximum of 6 mm. Finally, we carried out strict quality control to only select the subjects with mean head translation motion < 1 mm and mean head rotation < 1°.

The NeuroMark toolbox, available at http://www.yuhuidu.com/, was used to generate a common brain mask for ICA. First, using the volume at the first time point, a brain mask for each subject was calculated by setting voxels showing values > 90% of the mean value in the whole brain to 1. Then, a group mask was yielded by setting voxels included in > 90% of the individual masks to 1. Third, the correlations between the group mask and the individual mask for each subject were calculated. The correlations were calculated using voxels within the top 10 slices of the mask, within the bottom 10 slices of the mask, and within the whole mask, resulting in three correlation values for each subject. If a subject had three correlations greater than the

specified thresholds (0.75, 0.55, and 0.80), we included the subject for further fMRI analysis. Finally, the common brain mask was computed based on the selected subjects' masks.

Through this processing, the remaining subjects were divided into two groups (Group 1 and Group 2), each group containing 975 subjects. Between the two groups, there were no significant differences in age ($P = 0.9776$) and head motion (translation: $P = 0.7171$, rotation: $P = 0.5979$) using the two-sample *t*-test, and no significant differences in gender ($P = 0.9636$) using the χ^2 test. In addition, for all subjects in both groups, the fMRI acquisition was identical (repetition time: 0.7350 s, slice number: 64, slice size: 88×88, time point: 490).

(2) Functional Network Extraction Using SMART ICA Based on Real fMRI Data

We applied SMART ICA separately to the data of Group 1 and Group 2 to evaluate the reproducibility of results for both the group-level and subject-specific FNs. According to previous research [46, 47], small-scale FNs, each of which often includes many spatially remote brain regions, can be extracted using a small number of components, while large-scale FNs, each of which tends to be a small brain region, can be obtained using a relatively large number of components. Therefore, we divided the model order range into low (model orders = 20, 25, 30, and 35) and high (model orders = 85, 90, 95, and 100) numbers for separate validation. In the following, we take the low-model-order range as the instance for the explanation. In step 1, ICA was applied to the fMRI data to obtain initial group-level ICs for each given *N* (20, 25, 30, or 35), resulting in a total of 110 initial group-level ICs. In step 2, our clustering method was carried out to cluster 110 initial group-level ICs to obtain reliable group-level ICs. Since reliable group-level ICs not only consist of meaningful FNs but also involve some meaningless artifact ICs, in step 3, artifact ICs were removed from the reliable group-level ICs, and the remaining ICs were regarded as reliable group-level FNs. After that, we provided the linkage information between different-scale initial group-level ICs for each remaining cluster. In step 4, by taking the reliable group-level FNs as guidance, GIG-ICA was used to estimate the subject-specific FNs.

(3) Evaluation of SMART ICA Based on Results of Real fMRI Data

Based on the real fMRI data, in addition to the evaluation of the validity of the proposed clustering method and the ability to capture the linkage relationships between initial group-level ICs, we also evaluated the reproducibility of results obtained from Group 1 and Group 2. Specifically, for both low-model-order and high-model-order, we assessed whether the reliable group-level FNs from the two groups were similar in quantity and quality. Here, two groups of reliable group-level FNs were matched by performing greedy spatial correlation analysis using the Pearson correlation coefficient [14].

Using real fMRI data, we were also interested in assessing whether the FN correspondence and specificity were well preserved. For this goal, we used t-distributed stochastic neighbor embedding (t-SNE) [48–50] for the projection of all networks of all subjects into a two-dimensional space. To further investigate whether the resulting FNs can capture subject specificity well, we investigated the group differences in the subject-specific FNs between different sets at different ages. For each group (Group 1 or Group 2), all the subjects were divided into 5 sets: 45–47 (benchmark), 48–49, 50–51, 52–53, and 54–55 years. Subsequently, for each FN, we evaluated the group differences between $Set_{benchmark}$ and Set_{other} (e.g., Set_{48-49}) for each important voxel using a two-sample *t*-test. Here, the voxels that passed a right-tailed one-sample *t*-test ($P < 0.05$ with Bonferroni correction) based on all subjects were regarded as important voxels. We then summarized the *T*-values with $P < 0.05$ in two-sample *t*-tests and visualized the differences (for *T*-value > 0 and *T*-value < 0 separately) between different age subgroups of all FNs by using boxplots.

(4) Constructing Brain Functional Network Templates

Standardized brain functional templates are recognized for their utility in facilitating large-sample analyses and enhancing the robustness of findings. Here, we provide FN templates of both small- and large-scale to promote the unified and standardized analysis of fMRI data. Specifically, for both the low-model-order and the high-model-order ranges, we matched the reliable group-level FNs from Group 1 and Group 2 using a greedy spatial correlation analysis [14] according to the Pearson correlation coefficient. Pairs of FNs, where the Pearson correlation coefficient exceeded 0.5, were averaged to construct one FN template. For each template, we associated it with the Automated Anatomical Atlas 3 (AAL3) [51] using the Intelligent Analysis of Brain Connectivity (IABC) toolbox available at http://yuhuidu.com/. This process allowed us to identify the primary brain regions associated with each FN. The relevant brain regions were identified by calculating the overlap between the brain regions of AAL3 and the activated regions in the developed templates.

Code Availability

The code of SMART ICA is integrated into the toolbox IABC, which is accessible at http://www.yuhuidu.com/.

Result

Results from Simulated Data

(1) Results Based on Two Groups of Simulated Data with Group-common and Group-unique SMs

With the two groups of simulated data with group-common and group-unique SMs, we applied ICA using various model orders, resulting in 54 initial group-level ICs. Then, a graph constructed based on those initial group-level ICs was divided into 10 trees representing 10 clusters, of which the cluster centers were regarded as the reliable group-level ICs. In our results, all 10 reliable group-level ICs were retained as reliable group-level FNs, suggesting that the number of estimated reliable group-level FNs was the same as the number of real SM templates (six common and four unique templates) in the two groups.

Furthermore, the low correlation coefficients (Fig. 2A) among the 10 reliable group-level FNs indicate that each FN has a unique pattern. Fig. 2B shows the correlation matrix of the related 54 initial group-level ICs that are sorted according to cluster labels. The high intra-cluster similarity and

Fig. 2 The correlation matrix between reliable group-level FNs and the correlation matrix between initial group-level ICs based on simulated data with group-common and group-unique SMs. **A** Correlation matrix of 10 reliable group-level FNs. **B** Correlation matrix of 54 initial non-artifact group-level ICs after sorting according to the cluster labels. The black lines represent the division of the different clusters.
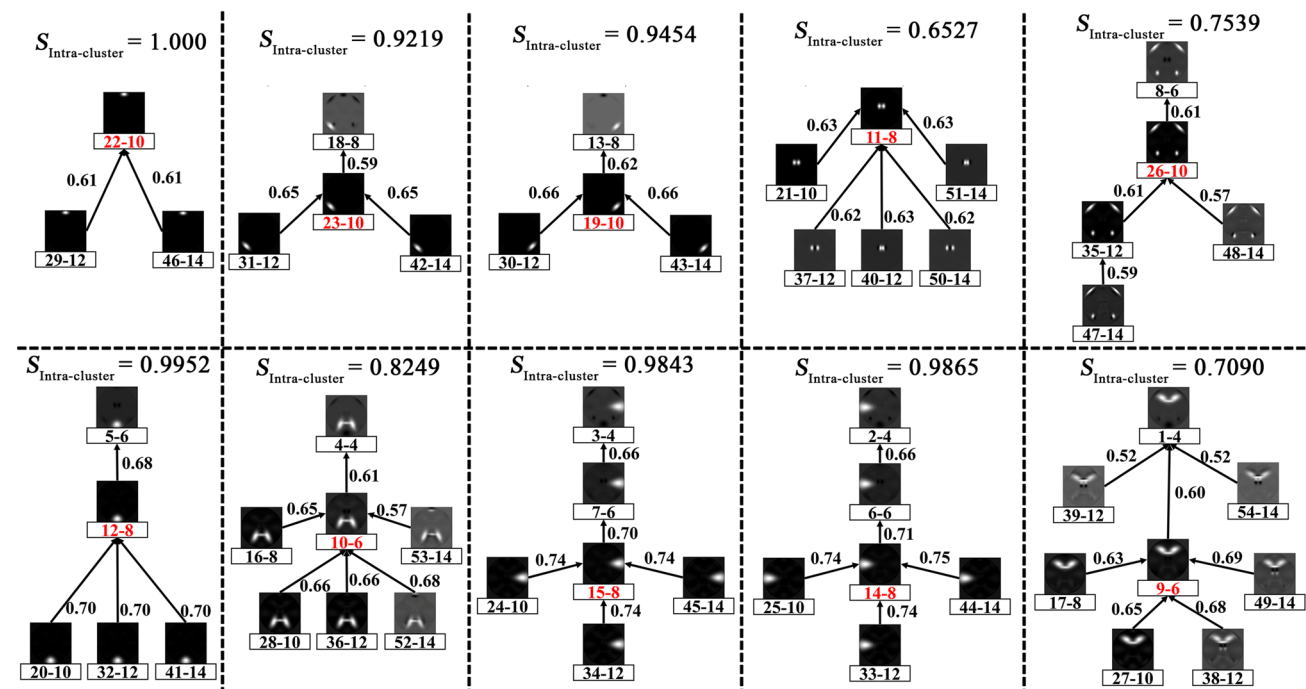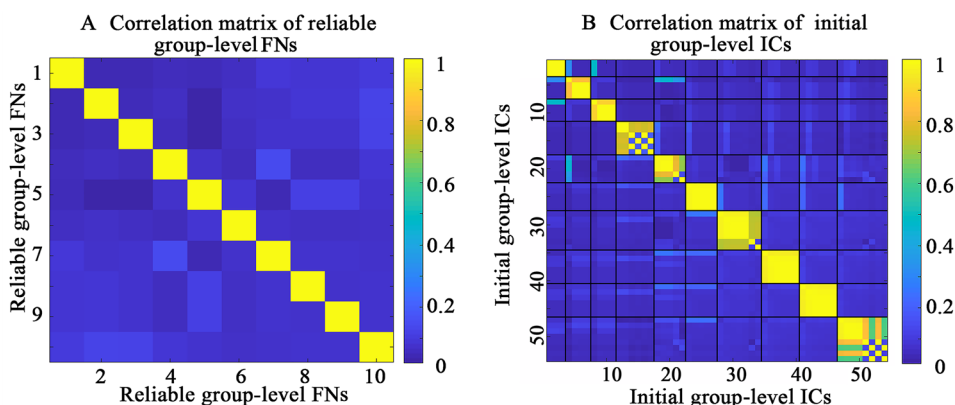

Fig. 3 Tree structures of the resulting clusters corresponding to 10 reliable group-level FNs. In each subfigure, $S_{\text{Intra-cluster}}$ represents the intra-cluster similarity. The 2D images show the initial group-level ICs. Below each 2D image, there is Number 1-Number 2, where Number 1 represents the ID of the IC in all initial group-level ICs, and Number 2 represents the model order of the IC. For the reliable group-level FN in each cluster, we display the Number 1-Number 2 in red. The 10 trees in this figure correspond to the 10 corresponding clusters in Fig. 2B.

low inter-cluster similarity provide strong evidence for the effectiveness of the proposed clustering method.

As summarized in Table S1, among the 10 reliable group-level FNs, 2, 4, and 4 FNs came from the ICA runs of which the model order was 6, 8, and 10, respectively, meaning that multi-model-order results were jointly used in our method. In Fig. 3, we present linkage information between different-scale ICs for each cluster, including intra-cluster similarity, the initial group-level ICs, the IDs of the ICs, the model orders corresponding to the ICs, and the following degree between ICs to depict the linkage relationships within each cluster. Our findings indicate that our method effectively captured the linkage relationships among results from different model orders.

The subject-specific FNs/TCs were computed based on the reliable group-level FNs, resulting in 8 FNs for each subject by the match with real SMs. Fig. 4 illustrates that the similarities between subject-specific FNs/TCs and real SMs/TCs across all subjects are greater than 0.9. Moreover, we show the FNs and TCs for one example subject of each group in Fig. S1. Taken together, we found that both the common and unique SMs were perfectly extracted and the similarities between the estimated FNs/TCs and real SMs/TCs were high for all subjects, which supports the conclusion that the proposed method-SMART ICA can effectively and accurately identify subject-specific FNs for the simulated data.

(2) Results Based on Six Groups of Simulated Data with Different Numbers of SMs

Based on six groups of simulated data with different numbers of SMs (numbers of SMs: 4, 6, 8, 10, 12, and 14), our experimental results showed that our method effectively extracts reliable group-level FNs and subject-specific FNs that are highly similar to the true SMs for all six groups. The numbers of reliable group-level ICs and reliable group-level FNs for six groups of simulated data are listed in Table S2. Fig. S2 illustrates that all mean similarities between subject-specific FNs/TCs and real SMs/TCs for all subjects across

all six groups are greater than 0.95. The above results support the conclusion that the SMART ICA is insensitive to the true number of SMs and the model order range setting.

Results from fMRI Data

By performing experiments using the real fMRI data, we obtained the initial group-level ICs, the reliable group-level FNs, and the subject-specific FNs for both Group 1 and Group 2 as well as for both the low-model-order and the high-model-order.

For Group 1, under the low-model-order range, 32 reliable group-level ICs were obtained by clustering 110 initial group-level ICs using our method. Subsequent removal of artifact ICs resulted in 24 reliable group-level FNs, with 76 corresponding to initial group-level ICs. Similarly, for Group 2, 33 reliable group-level ICs were obtained first, and then 25 reliable group-level FNs and 74 corresponding initial group-level ICs were retained after removal of artifact ICs. Regarding the high-model-order range, 102 and 99 reliable group-level ICs were obtained based on 370 initial group-level ICs for Group 1 and Group 2, respectively. After removing the group-level artifact ICs, 74 group-level FNs and 269 corresponding initial group-level ICs were retained for Group 1, and 69 group-level FNs and 268 corresponding initial group-level ICs were retained for Group 2. In summary, the results for both Group 1 and Group 2 exhibited remarkable consistency in terms of the quantity of reliable group-level FNs for both low-model-order and high-model-order conditions.

Regarding the quality of FNs, Fig. 5A, C reflect the low correlation coefficients between reliable group-level FNs of Group 1 and Group 2 under the low-model-order range, indicating the uniqueness of each reliable group-level FN. Furthermore, Fig. 5B, D show the correlation coefficient of initial group-level ICs corresponding to reliable group-level
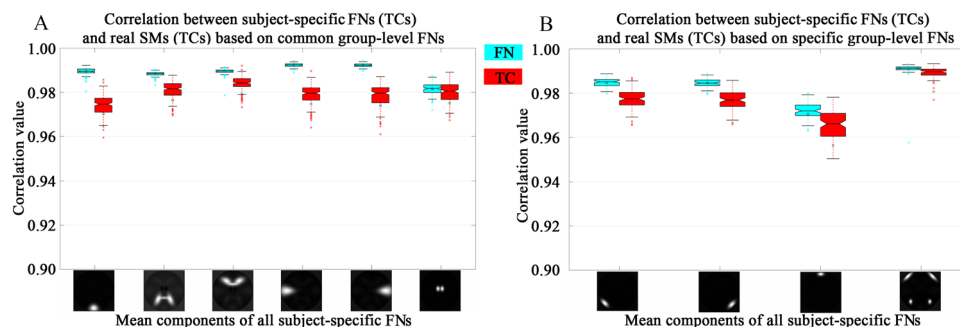

Fig. 4 The similarities reflected by the correlations between subject-specific FNs (TCs) and real SMs (TCs) across all subjects. The blue and red boxplots display the similarity results of FNs and TCs, respectively, with each boxplot showing the accuracy values of one

FN or TC across all subjects. Since both group-common and group-unique SMs were simulated, we show the accuracy results corresponding to the group-common SMs in subfigure **A** and the results corresponding to the group-unique SMs in subfigure **B**.

914

Neurosci. Bull. July, 2024, 40(7):905–920

Fig. 5 The correlation of reliable group-level FNs and initial group-level ICs based on real fMRI data for the low-model-order. **A** Correlation matrix of 24 reliable group-level FNs of Group 1. **B** Correlation matrix of 76 initial group-level ICs after sorting according to the cluster labels of Group 1. **C** Correlation matrix of 25 reliable group-level FNs of Group 2. **D** Correlation matrix of 74 initial group-level ICs after sorting according to the cluster labels of Group 2.
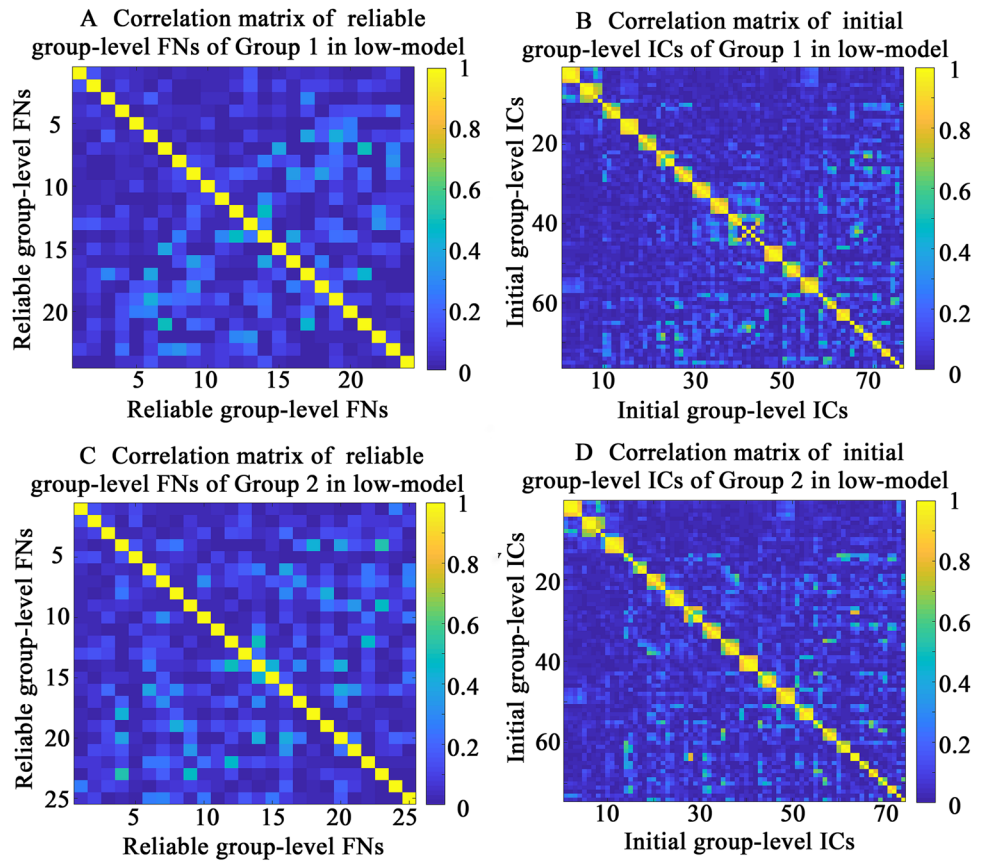

Fig. 6 The matched reliable group-level FNs of Group 1 and Group 2 for the low-model-order. In each subfigure, along with two matched FNs from the two groups, the similarity (Pearson correlation coefficient) between them is provided in R. In $R_{(Number\ 1,\ Number\ 2)}$, Number 1 and Number 2 represent the FN's ID in Group 1 and Group 2, respectively. All FNs are displayed after applying the Z-score transformation in this paper. The color bar represents the Z-score of each functional network.
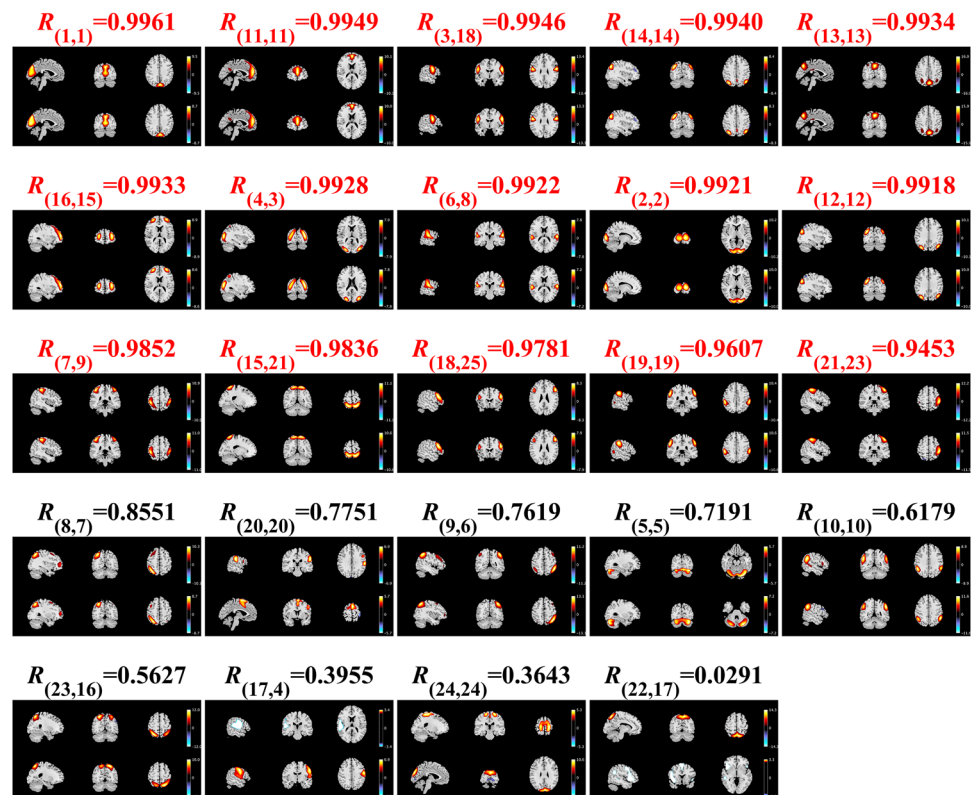
FNs for Group 1 and Group 2, respectively, supporting low inter-cluster similarities and high intra-cluster similarities for the clustering results. Due to the limited space, we demonstrate the relevant correlation results under the high-model-order range in Fig. S3. Overall, Figs. 5 and S3 support the effectiveness of the proposed method in clustering the initial group-level ICs based on real fMRI data.

To validate the robustness of our proposed method and the reproducibility of the results, we further compared the quality of the reliable group-level FNs between two groups (Group 1 and Group 2) for both the low-model-order and high-model-order. Fig. 6 and S4 show the matched results of the reliable group-level FNs for the low-model-order and high-model-order, respectively, which illustrates that the spatial similarities between Group 1 and Group 2 exceeded 0.9 for most of the reliable group-level FNs in both model ranges. Based on the above results, it can be concluded that the reliable group-level FNs extracted from two groups (with similar characteristics) were highly similar in quantity and quality, meaning strong robustness and high reproducibility.

One of the advantages of our method is its ability to jointly utilize multi-model-order results. For the low-model-order, among the 24 reliable group-level FNs of Group 1, 7, 7, 2, and 8 FNs were retained from the results of $k = 20, 25, 30$, and 35; among the 25 reliable group-level FNs of Group 2, 6, 6, 4, and 9 FNs were retained from the results of $k = 20, 25, 30$, and 35. For the high-model-order, among the 74 reliable group-level FNs of Group 1, 34, 17, 11, and 12 FNs were retained from the results of $k = 85, 90, 95$, and 100; among the 69 reliable group-level FNs of Group 2, 33, 20, 10, and 6 FNs were retained from the results of $k = 85, 90, 95$, and 100. Therefore, it seemed that the distribution of reliable ICs of Group 1 and Group 2 resemble each other, as summarized in Table S1, which supports the conclusion that our method can simultaneously take advantage of ICA results with different model order settings.

For the reliable group-level FNs in Group 1 and Group 2, which exhibited matched similarities greater than 0.9, the detailed linkage relationships (represented by the intra-cluster similarity, the initial group-level ICs, the IDs of the ICs, the model orders corresponding to the ICs, and the following degree between ICs) of the low-model-order are shown in Fig. 7. As there were many FNs (i.e., 53) with matched similarities > 0.9 between Group 1 and Group 2 for the high-model-order, we selected only the top 10 reliable group-level FNs with the highest between-group similarity to display their linkage relationships in the form of trees in Fig. S5. We found that, for the initial group-level ICs under a single model order setting, no more than one initial group-level IC was included in the same tree for almost all trees. Furthermore, the IC on a small scale (e.g., obtained using the number of components as 25) tended to be the leader to be followed by the IC on a bigger scale (e.g., obtained using the number of components as 30). To sum up, our method

is capable of capturing the linkage relationships among ICs in different scales.

We further visualized all subject-specific FNs using t-SNE, as depicted in Fig. 8A, B for the low-model-order range and Fig. S6A, B for the high-model-order range. These visualizations reveal that subject-specific FNs corresponding to the same reliable group-level FN are closely located but also exhibit subject-specific variability. The results support that both the correspondence and specificity of the subject-specific FNs are well guaranteed by using our method. Furthermore, as outlined in the methodology section, we conducted a comprehensive investigation into the age effect on subject-specific FNs using statistical analysis. Fig. 8C, D show the *T*-value results under the low-model-order range for Group 1 and Group 2. The related results from the high-model-order range are shown in Fig. S6C, D. The results indicate that while the age gaps gradually increase, the differences in subject-specific FNs between different subject sets also increase, supporting the idea that the subject-specific FNs extracted by our method can identify subtle specificity. It is important to note that $T > 0$ indicates that the *Z*-score of subject-specific FNs of subjects in Set_{other} is lower than that in $Set_{benchmark}$. Conversely, when $T < 0$, it signifies that the *Z*-score of subject-specific FNs of subjects in Set_{other} is higher than that in $Set_{benchmark}$. Our results reveal that differences with $T > 0$ are more pronounced than those with $T < 0$, indicating that, in general, reduced connectivity may be more evident along with the increasing age of subjects. All the results underscore the conclusion that the extracted FNs inherit subject-specific characteristics.

More importantly, two groups of brain network templates were obtained in our work. The templates from the low-model-order and high-model-order ranges include 21 and 65 FNs, respectively. The visualizations for the templates are shown in Fig. S7A and B, respectively. Comprehensive details, including the primary brain region and peak coordinates of each FN in the templates, are summarized in Table S3. These templates are available at http://yuhuidu.com/ and http://trendscenter.org/data/, facilitating the advancement and utilization of FNs in the field of neuroscience.

Discussion

Developing an effective method for identifying reliable brain functional networks from fMRI data is greatly needed in the neuroscience field. ICA has been widely applied to FN analysis using fMRI data. However, obtaining reliable brain FNs using ICA under the condition of an unknown number of components is challenging [21, 27]. In this paper, we propose a method, named SMART ICA, which automatically clusters ICs from multi-model-order ICA on fMRI data
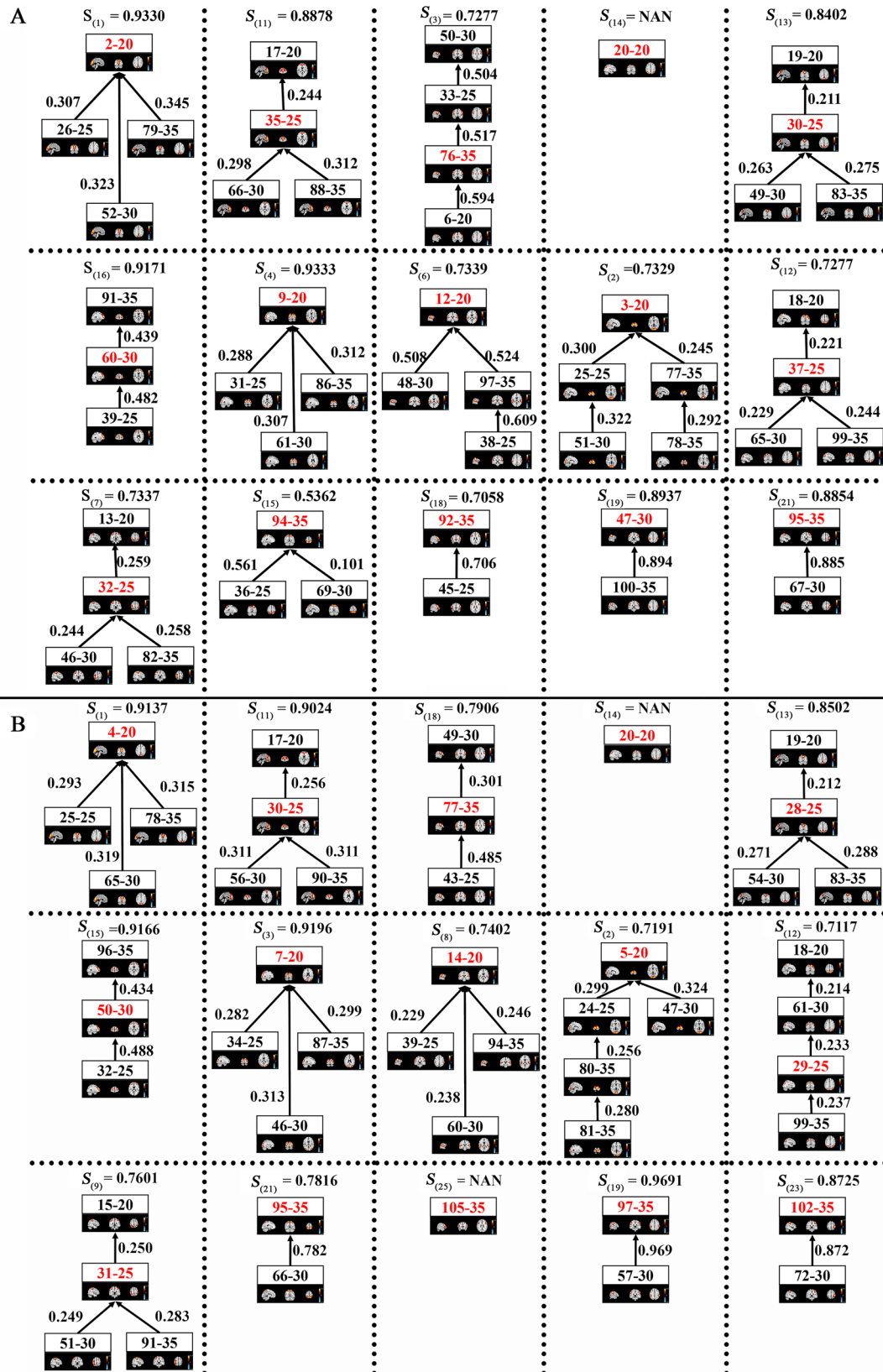
A

$S_{(1)} = 0.9330$
2-20
0.307 · 0.345
26-25 · 79-35
0.323
52-30

$S_{(11)} = 0.8878$
17-20
0.244
35-25
0.298 · 0.312
66-30 · 88-35

$S_{(3)} = 0.7277$
50-30
0.504
33-25
0.517
76-35
0.594
6-20

$S_{(14)} = $ NAN
20-20

$S_{(13)} = 0.8402$
19-20
0.211
30-25
0.263 · 0.275
49-30 · 83-35

$S_{(16)} = 0.9171$
91-35
0.439
60-30
0.482
39-25

$S_{(4)} = 0.9333$
9-20
0.288 · 0.312
31-25 · 86-35
0.307
61-30

$S_{(6)} = 0.7339$
12-20
0.508 · 0.524
48-30 · 97-35
0.609
38-25

$S_{(2)} = 0.7329$
3-20
0.300 · 0.245
25-25 · 77-35
0.322 · 0.292
51-30 · 78-35

$S_{(12)} = 0.7277$
18-20
0.221
37-25
0.229 · 0.244
65-30 · 99-35

$S_{(7)} = 0.7337$
13-20
0.259
32-25
0.244 · 0.258
46-30 · 82-35

$S_{(15)} = 0.5362$
94-35
0.561 · 0.101
36-25 · 69-30

$S_{(18)} = 0.7058$
92-35
0.706
45-25

$S_{(19)} = 0.8937$
47-30
0.894
100-35

$S_{(21)} = 0.8854$
95-35
0.885
67-30

B

$S_{(1)} = 0.9137$
4-20
0.293 · 0.315
25-25 · 78-35
0.319
65-30

$S_{(11)} = 0.9024$
17-20
0.256
30-25
0.311 · 0.311
56-30 · 90-35

$S_{(18)} = 0.7906$
49-30
0.301
77-35
0.485
43-25

$S_{(14)} = $ NAN
20-20

$S_{(13)} = 0.8502$
19-20
0.212
28-25
0.271 · 0.288
54-30 · 83-35

$S_{(15)} = 0.9166$
96-35
0.434
50-30
0.488
32-25

$S_{(3)} = 0.9196$
7-20
0.282 · 0.299
34-25 · 87-35
0.313
46-30

$S_{(8)} = 0.7402$
14-20
0.229 · 0.246
39-25 · 94-35
0.238
60-30

$S_{(2)} = 0.7191$
5-20
0.299 · 0.324
24-25 · 47-30
0.256
80-35
0.280
81-35

$S_{(12)} = 0.7117$
18-20
0.214
61-30
0.233
29-25
0.237
99-35

$S_{(9)} = 0.7601$
15-20
0.250
31-25
0.249 · 0.283
51-30 · 91-35

$S_{(21)} = 0.7816$
95-35
0.782
66-30

$S_{(25)} = $ NAN
105-35

$S_{(19)} = 0.9691$
97-35
0.969
57-30

$S_{(23)} = 0.8725$
102-35
0.872
72-30

◀**Fig. 7** The linkage relationships of initial group-level ICs within each cluster that have a high reproducibility between Group 1 and Group 2 with matched similarity > 0.9 for the low-model-order. **A** and **B** represent the results of Group 1 and Group 2, respectively. For each cluster, $S_{(Number\ 1)}$ represents the intra-cluster similarity and Number 1 represents the ID of the reliable group-level FN. The black boxes show the initial group-level ICs. Number 1 represents the ID of this IC in all initial group-level ICs, Number 2 represents the model order used, and Number 1-Number 2 in red corresponds to the reliable group-level FN in this cluster. The color bar represents the *Z*-score of each IC. If there is only one IC in the cluster, then the intra-cluster similarity cannot be calculated, so it is represented by the null value "NAN".

to obtain reliable FNs, which avoids setting a specific ICA model order. Importantly, our method provides direct linkage information among FNs deduced from different model orders by simplifying the graph that reflects the relationship among those ICs into a tree structure.

The automatic extraction of reliable FNs is the most important advantage of SMART ICA. In the method, effective clustering is achieved on ICs from multiple model orders using our proposed splitting and merging clustering method combined with a graph simplification technique. Our method utilizes the intra-cluster distance of each cluster, the inter-cluster distance of two clusters, and the mean inter-cluster distance to measure within-cluster quality and inter-cluster separability to guide the split and merge process. By iteratively performing split and merge operations, automatic and robust clustering for the integration of FN information across varying model orders is achieved without the need for specifying an ICA model order. Although setting a range of model order is needed for the multi-model-order analysis in our method, it is much easier to set a range compared to determine the optimal one in the traditional ICA analysis.

Based on two groups of simulated data with common and unique SMs and six groups of simulated data with different numbers of SMs, the number of reliable FNs is consistent with the number of real SMs. The high similarities between the subject-specific FNs and real SMs support the effectiveness of SMART ICA and the reliability of the extracted FNs. Based on two groups of simulated data that have both
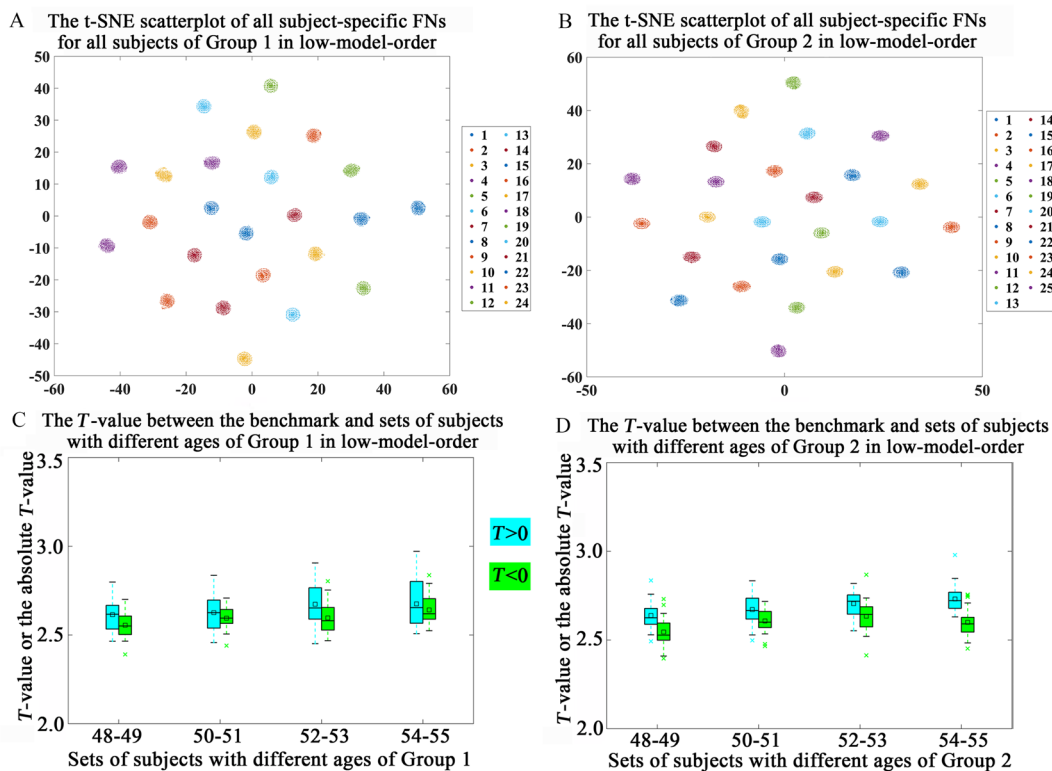

Fig. 8 Correspondence and specificity of subject-specific FNs for the low-model-order. **A** and **B** display the projections of subject-specific FNs of Group 1 and Group 2 using t-SNE, respectively. In these scatterplots, the FNs corresponding to the same reliable group-level FN are represented using the same color. **C** and **D** display the differences in subject-specific FNs between the $Set_{benchmark}$ and the Set_{other} with different ages of Group 1 and Group 2, respectively. The difference of one FN between $Set_{benchmark}$ and Set_{other} is represented by the mean *T*-value with $P < 0.05$ across all voxels within the important voxels. The *x*-axis of **C** and **D** represents the sets of subjects of different ages. The *y*-axis of **C** and **D** represent the absolute *T*-value obtained by two-sample *t*-tests.

common and unique SMs, the subject-specific FNs extracted by SMART ICA are highly similar to the real SMs, indicating the ability of our method to capture subject variability. By applying SMART ICA with the same parameter setting to six groups of simulated data with different numbers of SMs, the resulting subject-specific FNs are also highly similar to the real SMs. The results indicate that SMART ICA is insensitive to parameters and easy to use.

For the real fMRI data, the FNs show high similarities in the spatial patterns between two age-matched cohorts for both the small and large scales. In addition, the quantity of reliable group-level FNs between the two groups is very close. More interestingly, we also found that the difference in subject-specific FNs between different age groups showed an increasing trend as the age gap increased. Moreover, our results revealed that the reduced connectivity along with aging seems to be more pronounced than the increased connectivity. Many researchers [52–54] have found that, as age increases, the strength of functional connectivity in the human brain tends to decrease, which supports our findings. Taken together, our results demonstrate that the FNs obtained by SMART ICA can inherit the subject specificity, so it is feasible to study and discover well-characterized biomarkers from subject-specific FNs using the SMART ICA method.

The exploration of the linkage relationship between different-scale FNs (corresponding to different model orders) is another important strength of our method. Using the graph simplification technique, the linkage provides direct information about the following and leading (or being followed) relationships between the ICs within the same cluster. Based on the simplified tree structure with sparse edges, it is more convincing to select a dominant IC as the representative and reliable IC for further analysis. As expected, the ICs that are grouped into one cluster show a high spatial similarity. That means the ICs having the same biological meaning tend to be closely linked and thus grouped. We also found that for some well-known stable FNs such as the default mode network, visual network, and motor-related networks, each of those clusters seems to include more ICs, compared to the less important FNs or artifact-related ICs. We think one possible reason is that important FNs can always be successfully extracted regardless of model orders.

In summary, we propose a novel brain FN extraction method, called SMART ICA, which automatically clusters ICs obtained from multi-model-order ICA to output reliable and accurate brain FNs. Furthermore, SMART ICA enables the exploration of complex relationships among networks under different scales, which may be beneficial for understanding the mechanism of brain FNs and how they collaboratively or independently operate. In addition, reliable templates for small-scale and large-scale are provided based on fMRI data of 1,950 subjects, which furnish a benchmark for further study of individual functional networks. Collectively, the SMART ICA method holds promise for advancing the application of ICA in fMRI analysis.

References

1. Zhao B, Li T, Smith SM, Xiong D, Wang X, Yang Y. Common variants contribute to intrinsic human brain functional networks. Nat Genet 2022, 54: 508–517.
2. Zanin M. Simplifying functional network representation and interpretation through causality clustering. Sci Rep 2021, 11: 15378.
3. Chen Y, Tang JH, de Stefano LA, Wenger MJ, Ding L, Craft MA, *et al*. Electrophysiological resting state brain network and episodic memory in healthy aging adults. Neuroimage 2022, 253: 118926.
4. Chong JSX, Ng KK, Tandi J, Wang C, Poh JH, Lo JC, *et al*. Longitudinal changes in the cerebral cortex functional organization of healthy elderly. J Neurosci 2019, 39: 5534–5550.
5. Váša F, Romero-Garcia R, Kitzbichler MG, Seidlitz J, Whitaker KJ, Vaghi MM, *et al*. Conservative and disruptive modes of adolescent change in human brain functional connectivity. Proc Natl Acad Sci U S A 2020, 117: 3248–3253.
6. Du Y, He X, Kochunov P, Pearlson G, Hong LE, van Erp TGM, *et al*. A new multimodality fusion classification approach to explore the uniqueness of schizophrenia and autism spectrum disorder. Hum Brain Mapp 2022, 43: 3887–3903.
7. Qian X, Castellanos FX, Uddin LQ, Loo BRY, Liu S, Koh HL, *et al*. Large-scale brain functional network topology disruptions underlie symptom heterogeneity in children with attention-deficit/hyperactivity disorder. Neuroimage Clin 2019, 21: 101600.
8. Khatri U, Kwon GR. Alzheimer's disease diagnosis and biomarker analysis using resting-state functional MRI functional brain network with multi-measures features and hippocampal subfield and amygdala volume of structural MRI. Front Aging Neurosci 2022, 14: 818871.
9. Li X, Zhang K, He X, Zhou J, Jin C, Shen L, *et al*. Structural, functional, and molecular imaging of autism spectrum disorder. Neurosci Bull 2021, 37: 1051–1071.
10. Yang J, Gohel S, Vachha B. Current methods and new directions in resting state fMRI. Clin Imaging 2020, 65: 47–53.
11. Wang Z, Xin J, Wang Z, Yao Y, Zhao Y, Qian W. Brain functional network modeling and analysis based on fMRI: A systematic review. Cogn Neurodyn 2021, 15: 389–403.
12. Patil AU, Madathil D, Huang CM. Healthy aging alters the functional connectivity of creative cognition in the default mode network and cerebellar network. Front Aging Neurosci 2021, 13: 607988.

13. Du Y, Guo Y, Calhoun VD. Aging brain shows joint declines in brain within-network connectivity and between-network connectivity: A large-sample study (*N* > 6, 000). Front Aging Neurosci 2023, 15: 1159054.

14. Du Y, Fu Z, Sui J, Gao S, Xing Y, Lin D, *et al*. NeuroMark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders. Neuroimage Clin 2020, 28: 102375.

15. Bi XA, Zhao J, Xu Q, Sun Q, Wang Z. Abnormal functional connectivity of resting state network detection based on linear ICA analysis in autism spectrum disorder. Front Physiol 2018, 9: 475.

16. Rabany L, Brocke S, Calhoun VD, Pittman B, Corbera S, Wexler BE, *et al*. Dynamic functional connectivity in schizophrenia and autism spectrum disorder: Convergence, divergence and classification. Neuroimage Clin 2019, 24: 101966.

17. Rai S, Griffiths KR, Breukelaar IA, Barreiros AR, Chen W, Boyce P, *et al*. Default-mode and fronto-parietal network connectivity during rest distinguishes asymptomatic patients with bipolar disorder and major depressive disorder. Transl Psychiatry 2021, 11: 547.

18. Gu Y, Li L, Zhang Y, Ma J, Yang C, Xiao Y, *et al*. The overlapping modular organization of human brain functional networks across the adult lifespan. Neuroimage 2022, 253: 119125.

19. Xu J, Calhoun VD, Worhunsky PD, Xiang H, Li J, Wall JT, *et al*. Functional network overlap as revealed by fMRI using sICA and its potential relationships with functional heterogeneity, balanced excitation and inhibition, and sparseness of neuron activity. PLoS One 2015, 10: e0117029.

20. Li YO, Adali T, Calhoun VD. Estimating the number of independent components for functional magnetic resonance imaging data. Hum Brain Mapp 2007, 28: 1251–1266.

21. Fu GS, Anderson M, Adalı T. Likelihood estimators for dependent samples and their application to order detection. IEEE Trans Signal Process 2014, 62: 4237–4244.

22. Cordes D, Nandy RR. Estimation of the intrinsic dimensionality of fMRI data. Neuroimage 2006, 29: 145–154.

23. Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging 2004, 23: 137–152.

24. Seghouane AK, Shokouhi N. Consistent estimation of dimensionality for data-driven methods in fMRI analysis. IEEE Trans Med Imaging 2019, 38: 493–503.

25. Hui M, Li J, Wen X, Yao L, Long Z. An empirical comparison of information-theoretic criteria in estimating the number of independent components of fMRI data. PLoS One 2011, 6: e29274.

26. Kuang LD, Lin QH, Gong XF, Cong F, Sui J, Calhoun VD. Model order effects on ICA of resting-state complex-valued fMRI data: Application to schizophrenia. J Neurosci Methods 2018, 304: 24–38.

27. Hu G, Waters AB, Aslan S, Frederick B, Cong F, Nickerson LD. Snowball ICA: A model order free independent component analysis strategy for functional magnetic resonance imaging data. Front Neurosci 2020, 14: 569657.

28. Meng X, Iraji A, Fu Z, Kochunov P, Belger A, Ford J, *et al*. Multimodel order independent component analysis: A data-driven method for evaluating brain functional network connectivity within and between multiple spatial scales. Brain Connect 2022, 12: 617–628.

29. Calhoun VD, Liu J, Adali T. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. Neuroimage 2009, 45: S163–S172.

30. Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A method for making group inferences from functional MRI data using independent component analysis. Hum Brain Mapp 2001, 14: 140–151.

31. Du Y, Fan Y. Group information guided ICA for fMRI data analysis. Neuroimage 2013, 69: 157–197.

32. Amari S, Cichocki A, Yang HH. A new learning algorithm for blind signal separation. In: Neural Information Processing Systems. Denver: ACM, 1995: 757–763.

33. Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. Neural Comput 1995, 7: 1129–1159.

34. Himberg J, Hyvarinen A. Icasso: software for investigating the reliability of ICA estimates by clustering and visualization. 2003 IEEE XIII Workshop on Neural Networks for Signal Processing. Toulouse, France. IEEE, 2003: 259–268.

35. Bai L, Liang J, Du H, Guo Y. A novel community detection algorithm based on simplification of complex networks. Knowl Based Syst 2018, 143: 58–64.

36. Lei J, Jiang T, Wu K, Du H, Zhu G, Wang Z. Robust K-means algorithm with automatically splitting and merging clusters and its applications for surveillance data. Multimed Tools Appl 2016, 75: 12043–12059.

37. Matas J, Chum O, Urban M, Pajdla T. Robust wide-baseline stereo from maximally stable extremal regions. Image Vis Comput 2004, 22: 761–767.

38. Khosravi MH, Hassanpour H. Image quality assessment using a novel region smoothness measure. J Vis Commun Image Represent 2019, 60: 217–228.

39. Nistér D, Stewénius H. Linear time maximally stable extremal regions. In: Computer Vision–ECCV 2008. Berlin: Springer, 2008: 183–196.

40. Du Y, Allen EA, He H, Sui J, Wu L, Calhoun VD. Artifact removal in the context of group ICA: A comparison of single-subject and group approaches. Hum Brain Mapp 2016, 37: 1005–1025.

41. Griffanti L, Douaud G, Bijsterbosch J, Evangelisti S, Alfaro-Almagro F, Glasser MF, *et al*. Hand classification of fMRI ICA noise components. Neuroimage 2017, 154: 188–205.

42. Beckmann C, MacKay C, Filippini N, Smith S. Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. Neuroimage 2009, 47: S148.

43. Kam TE, Zhang H, Shen D. A novel deep learning framework on brain functional networks for early MCI diagnosis. MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part III 11 2018: 293–301.

44. Erhardt EB, Allen EA, Wei Y, Eichele T, Calhoun VD. SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability. Neuroimage 2012, 59: 4160–4167.

45. Allen NE, Sudlow C, Peakman T, Collins R, UK Biobank. UK biobank data: Come and get it. Sci Transl Med 2014, 6: 2244.

46. Wei P, Bao R, Fan Y. Comparing the reliability of different ICA algorithms for fMRI analysis. PLoS One 2022, 17: e0270556.

47. Abou-Elseoud A, Starck T, Remes J, Nikkinen J, Tervonen O, Kiviniemi V. The effect of model order selection in group PICA. Hum Brain Mapp 2010, 31: 1207–1216.

48. Hinton GE, Roweis S. Stochastic neighbor embedding. In: Neural Information Processing Systems 2002: 857–864.

49. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008, 9: 2579–2605.

50. Jacobs RA. Increased rates of convergence through learning rate adaptation. Neural Netw 1988, 1: 295–307.

51. Rolls ET, Huang CC, Lin CP, Feng J, Joliot M. Automated anatomical labelling atlas 3. Neuroimage 2020, 206: 116189.

52. Varangis E, Habeck CG, Razlighi QR, Stern Y. The effect of aging on resting state connectivity of predefined networks in the brain. Front Aging Neurosci 2019, 11: 234.

53. Geerligs L, Maurits NM, Renken RJ, Lorist MM. Reduced specificity of functional connectivity in the aging brain during task performance. Hum Brain Mapp 2014, 35: 319–330.

54. Wang L, Li Y, Metzak P, He Y, Woodward TS. Age-related changes in topological patterns of large-scale brain functional networks during memory encoding and recognition. Neuroimage 2010, 50: 862–872.