RESEARCH

State-transition dynamics of resting-state functional magnetic resonance imaging data: model comparison and test-to-retest analysis

Saiful Islam[2], Pitambar Khanra[1], Johan Nakuci[3], Sarah F. Muldoon[1,2,4], Takamitsu Watanabe[5] and Naoki Masuda[1,2*]

Abstract

Electroencephalogram (EEG) microstate analysis entails finding dynamics of quasi-stable and generally recurrent discrete states in multichannel EEG time series data and relating properties of the estimated state-transition dynamics to observables such as cognition and behavior. While microstate analysis has been widely employed to analyze EEG data, its use remains less prevalent in functional magnetic resonance imaging (fMRI) data, largely due to the slower timescale of such data. In the present study, we extend various data clustering methods used in EEG microstate analysis to resting-state fMRI data from healthy humans to extract their state-transition dynamics. We show that the quality of clustering is on par with that for various microstate analyses of EEG data. We then develop a method for examining test–retest reliability of the discrete-state transition dynamics between fMRI sessions and show that the within-participant test–retest reliability is higher than between-participant test–retest reliability for different indices of state-transition dynamics, different networks, and different data sets. This result suggests that state-transition dynamics analysis of fMRI data could discriminate between different individuals and is a promising tool for performing fingerprinting analysis of individuals.

Keywords fMRI, EEG, MEG, Microstates, Clustering, Dynamics, Test–retest reliability, Fingerprinting

*Correspondence:
Naoki Masuda
naokimas@buffalo.edu
[1] Department of Mathematics , University at Buffalo, State University of New York at Buffalo, 244 Mathematics Building , Buffalo 14260, NY, USA
[2] Institute for Artificial Intelligence and Data Science, University at Buffalo, State University of New York at Buffalo, 215 Lockwood Hall, Buffalo 14260, NY, USA
[3] School of Psychology, Georgia Institute of Technology, North Avenue, Atlanta 30332, GA, USA
[4] Neuroscience Program, University at Buffalo, State University of New York at Buffalo, 955 Main Street, Buffalo 14203, NY, USA
[5] International Research Centre for Neurointelligence, The University of Tokyo Institutes for Advanced Study, 731 Hongo Bunkyo-ku, Tokyo 113-0033, Japan

Introduction

Activity of the human brain is dynamic even at rest, and brain dynamics on various spatial scales are considered to drive myriad functions of the brain [1–4]. Multiple methods to characterize brain dynamics have been proposed, many of which rely on the detection of brain states and quantification of how the brain transitions through such states. Microstate analysis is an early-proposed method for estimating discrete states in electroencephalogram (EEG) data [5–7]. EEG microstate analysis usually entails clustering of multi-electrode EEG signals, with each data point to be clustered corresponding to a time point of the measurement. Each cluster, or microstate, is a representation of a global functional state of the brain. Microstates obtained from resting-state EEG data tend to last about 100 ms and are reproducible [6,

Islam *et al. BMC Neuroscience*   (2024) 25:14

Page 2 of 19

8–11]. Microstate analysis has been extended for magnetoencephalography (MEG) data, with the microstates being estimated by conventional clustering methods [12, 13] or the hidden-Markov model (HMM) [12, 14] among other methods. Microstate analysis in its original sense (i.e., detecting and utilizing microstates lasting about 100 ms) does not directly apply to functional magnetic resonance imaging (fMRI) data because the temporal resolution of fMRI is limited, preventing one from detecting dynamics on the timescale of 100 ms. One direction to resolve this limitation is to use EEG microstate analysis results to inform states in fMRI data [15–18]. An alternative approach is to estimate and use state-transition dynamics of spatial fMRI signals, as microstate analysis does for EEG (and MEG) data, regardless of different time resolutions between fMRI and EEG/MEG. Such state-transition dynamics for fMRI data have been estimated by data clustering algorithms as in the case of the EEG/MEG microstate analysis [19–22], the HMM or its variants [21–27], and energy landscape analysis [28–31]. Each discrete state in fMRI data corresponds to a vector of activity patterns at specified regions of interests (ROIs) [22, 32–35], or a functional network among ROIs [19–21, 23, 25, 36–39].

In general, successful single individual inferences from neuroimaging data would suggest their potential applications for both scientific investigations and clinical practice. Research has shown that functional networks from fMRI data can be used as a reliable fingerprint of human individuals through test–retest analyses [40–45]. Test–retest reliability has also been assessed for dynamic functional networks estimated from fMRI data [46–48], whereas test–retest reliability for dynamic functional networks has been reported to be lower than that for static functional networks [46, 47]. With this study, we are interested in test–retest reliability of state-transition dynamics in fMRI data, which has been underexplored.

In the present study, we assess the potential effectiveness of dynamics of discrete states estimated from fMRI data at fingerprinting individuals. Here, we use fMRI data as multivariate time series, each dimension of which represents a single ROI, akin to microstate analysis for EEG and MEG data. This approach contrasts with the aforementioned prior studies on test–retest reliability of dynamic functional networks. Our analysis involves examination of what methodological choices (e.g., the clustering method applied to the fMRI data to define discrete states, the number of clusters identified, and the indices used to characterize the estimated state transition dynamics) yield a higher test–retest reliability of the state-transition dynamics; such an assessment has previously been carried out for EEG microstate analysis [9]. Based on a permutation test to quantify test–retest reliability, we show that, in general, transitory dynamics of discrete states estimated for fMRI data yield higher within-participant than between-participant test–retest reliability across clustering methods, the number of clusters, observables of the state-transition dynamics, two sets of ROIs, and two data sets. Code for computing dynamics of discrete states and their test–retest reliability used in the present paper is available on GitHub [49].

Methods

Midnight Scan Club data

We use the resting-state fMRI data provided by the Midnight Scan Club (MSC) project [41]. The MSC's resting-state fMRI data consist of recording from ten healthy human adults [age: 29.1 ± 3.3 (average $\pm$ standard deviation); five males and five females] over ten consecutive nights. A single recording session of the resting-state fMRI experiment lasted for 30 mins, resulting in 818 volumes. The imaging was performed on a Siemens TRIO 3T MRI scanner. All functional imaging was performed using an echo planar imaging (EPI) sequence (TR = 2.2 s, TE = 27 ms, flip angle = 90°, voxel size = 4 mm × 4 mm × 4 mm, 36 slices).

It was originally reported that the eighth participant (i.e., MSC08) fell asleep, showed frequent and prolonged eye closures, and had systematically large head motion, yielding considerably less reliable data than those obtained from the other participants [41]. In our previous work, we also noticed that the quality of the data analysis fluctuated considerably more across the different sessions for the tenth participant (i.e., MSC10) than for the other participants except MSC08 [50]. Therefore, we excluded MSC08 and MSC10, and only used the remaining eight participants (age: 29.1 ± 3.4; four males and four females) in the following analysis.

We used SPM12 (http://www.fil.ion.ucl.ac.uk/spm) to preprocess the resting-state functional images. Specifically, we first conducted realignment, unwrapping, slice-timing correction, and normalization to a standard template (ICBM 152). Then, we performed regression analyses to remove the effects of head motion, white matter signals, and cerebrospinal fluid signals. Lastly, we conducted band-pass temporal filtering (0.01–0.1 Hz).

We used a DMN composed of 12 ROIs [51]. To optionally reduce the dimension of the DMN, which may improve the estimation of discrete states, we averaged over each pair of the symmetrically located right- and left-hemisphere ROIs into one observable. The symmetrized DMN has eight ROIs because four ROIs (i.e., amPFC, vmPFC, pCC, and retro splen) in the original coordinate system are approximately on the midline and therefore have not undergone the averaging over the right- and left-hemisphere ROIs [31].

Islam *et al. BMC Neuroscience*   (2024) 25:14

Page 3 of 19

In addition to the DMN, we also analyzed the so-called whole-brain network. We determined the regions of interest (ROIs) of the whole-brain network by employing the 264 spherical ROIs whose coordinates were identified in a previous study [52]. We then removed 50 ROIs labelled 'uncertain' or 'subcortical', resulting in 214 ROIs. We removed these 50 ROIs because they are located in the peripheral or border areas of the brain such that signals from these regions are easily contaminated with cerebrospinal fluid (CSF) signals. The 214 ROIs were labeled either of the following nine functionally different brain systems: auditory network, dorsal attention network (DAN), ventral attention network (VAN), cingulo-opercular network (CON), default mode network (DMN), fronto-parietal network (FPN), salience network (SAN), somatosensory and motor network (SMN), or visual network. We merged the DAN, VAN, and CON into an attention network (ATN) to reduce the number of observables from nine to seven, as we did in our previous studies [50, 53]. This is because the DAN, VAN, and CON have been suggested to be responsible for similar attention-related cognitive activity [52]. We then calculated the average fMRI signal for each of the seven systems by first averaging the signal over the volumes in the sphere of radius 4 mm centered around the provided coordinate of each ROI [52], and then averaging the signal over all ROIs belonging to the system (e.g., 13 ROIs in the auditory network). We call the thus obtained seven-dimensional system the whole-brain network. It should be noted that the DMN constitutes one ROI in the whole-brain network, whereas the DMN described above as a system of ROIs is composed of either 8 or 12 ROIs depending on whether or not we average over symmetrically located ROIs.

Human Connectome Project data

We also analyzed the fMRI data recorded from healthy human adults and shared as the S1200 data in the Human Connectome Project (HCP) [54]. In the S1200 data, 1200 adults between 22-35 years old underwent four sessions of 15-min EPI sequence with a 3T Siemens Connectome-Skyra (TR = 0.72 s, TE = 33.1 ms, 72 slices, 2.0 mm isotropic, field of view (FOV) = 208×180 mm) and a T1-weighted sequence (TR = 2.4 s, TE = 2.14 ms, 0.7 mm isotropic, FOV = 224×224 mm). To circumvent any potential influence of familial relationships, we use a subset of 100 participants that do not share any family relationship, called "100 unrelated participants" (age: 29.4 ± 3.5; 46 males and 54 females), released by the HCP. All these 100 participants completed both diffusion weighted MRI and two resting-state fMRI scans.

Each participant underwent two sessions of resting-state fMRI recording, and each session consisted of both Left-Right (LR) and Right-Left (RL) phases. In the following text, we refer to phases as sessions. Therefore, each participant's data consist of four sessions. We used data from participants with at least 1150 volumes in each of the four sessions after we had removed volumes with motion artifacts, resulting in a final analysis of 87 participants (age: 29.8 ± 3.4; 40 males and 47 females). For the 87 participants, we removed the volumes with motion artifacts and then used the last 1150 volumes in each session, with the aim of removing possible transient effects.

We employed independent component analysis (ICA) to remove nuisance and motion signals [55]. Then, any volumes with frame displacement greater than 0.2 mm [56] were excised [57]. This is because the ICA-FIX pipeline has been found not to fully remove motion-related artifacts [58, 59]. Next, we standardized each voxel by subtracting the temporal mean, and then global signal regression (see "Global signal removal" section) was carried out.

We averaged the fMRI signal over all the voxels within each ROI of the AAL atlas [60] in each volume. We remark that the AAL atlas is composed of 116 ROIs. In order to map these ROIs to representative brain systems, we first mapped each of the cortical ROIs to the parcellation scheme from the Schaefer-100 atlas [61]. We based the assignment of the ROI to the brain system that minimized the Euclidian distance from the centroid of an ROI in the AAL to the corresponding centroid of an ROI in the Schaefer atlas. We provide the correspondence between each AAL ROI and the brain system in Additional file 1: section S1. After we assigned each ROI to a system, we removed 42 ROIs labeled 'subcortical' or 'cerebellar', which yielded 74 ROIs. These 74 ROIs were then assigned to one of the $N = 7$ functionally different brain networks: control network, DMN, DAN, limbic network, salience/ventral attention network, somatomotor network, and visual network. We call this seven-dimensional system the whole-brain network for the HCP data. Similarly to the case of the whole-brain network for the MSC data, we first averaged the fMRI signal over the voxels within each ROI and then further averaged the signal over the ROIs belonging to the same system (e.g., 59 ROIs belonging to the DMN).

Global signal removal

We denote the fMRI time series for a session by $\{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$, where T is the number of volumes (i.e., time points), $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,\tilde{N}})$ is the fMRI signal at time t, and $\tilde{N}$ is the number of ROIs with which we compute the global signal. Note that $\tilde{N}$ may be larger than N, which occurs when we define a global signal widely from the brain including ROIs that we do not use for estimating

Islam *et al. BMC Neuroscience* (2024) 25:14

Page 4 of 19

the discrete states. The global signal is the average of the signal over all the $\tilde{N}$ ROIs at each time, i.e.,

$$\bar{x}_t = \frac{\sum_{i=1}^{\tilde{N}} x_{t,i}}{\tilde{N}}. \tag{1}$$

We remove the global signal [62] by subtracting $\bar{x}_t$ from each $x_{t,i}$ (with $i \in \{1, \ldots, \tilde{N}\}$) and dividing the result by the standard deviation, i.e.,

$$\sigma_t = \sqrt{\frac{\sum_{i=1}^{\tilde{N}} (x_{t,i} - \bar{x}_t)^2}{\tilde{N}}}. \tag{2}$$

We carry out this procedure for each t.

The global signal in resting-state fMRI data is considered to primarily consist of physiological noise stemming from different factors such as respiration, scanner-related artifacts, and motion-related artifacts. By removing the global signal, several quality-control metrics are improved, the anatomical specificity of functional-connectivity patterns is enhanced, and there is a potential increase in behavioral variance [63, 64].

It is a common practice to calculate the global signal using the gray matter, white matter, and CSF (e.g., [41, 51]). However, we calculate the global signal using the gray matter but not the white matter or CSF [59], as explained above, for the following reasons. First, white matter noise and CSF-oriented noise were removed in the preprocessing procedure that aims to suppress the effect of motion. Second, white matter and CSF signals contribute little to the fMRI signal in the gray matter. In fact, the characteristics of the blood-oxygen-level-dependent (BOLD) signal are substantially different in the gray matter compared with the white matter and even more so with CSF, which should not contain any signal associated with the BOLD signal. Therefore, including the white matter and CSF in the global signal calculation is more likely to induce distortions in the signal [55, 65].

For the DMN obtained from the MSC data, we first removed the global signal calculated over the $\tilde{N} = 30$ ROIs in the coordinate system provided by [51], which included the $N = 12$ ROIs in the DMN. Then, we compared three treatments of global signal removal for the DMN as follows. In the first and second treatments, we then removed the global signal calculated from the $\tilde{N} = 12$ DMN ROIs from each of the 12 ROIs in the DMN. Next, we averaged the obtained time series over each symmetric pair of DMN ROIs corresponding to the two hemispheres. If the ROI is roughly on the midline, there is no such symmetric pair of ROIs, in which case we only removed the global signal. After aggregating the symmetric pairs of ROIs in this manner, there are $N = 8$ ROIs in the DMN. This concludes the first treatment.

In the case of the second treatment, we additionally removed the global signal calculated over the $\tilde{N} = N = 8$ ROIs. In the third treatment, after the removal of the global signal calculated over 30 ROIs, which is common for all the three treatments, we further removed global signal calculated from the $\tilde{N} = 12$ DMN ROIs from each of the 12 ROIs. We do not further process the data. Therefore, with the third treatment, the final DMN consists of $N = 12$ ROIs.

For the whole-brain network obtained from the MSC data, we first removed the global signal computed from the $\tilde{N} = 264$ ROIs. Then, we extracted $N = 7$—dimensional time series as described in "Midnight Scan Club data" section. Finally, we further removed the global signal computed from the $\tilde{N} = N = 7$ ROIs in the whole-brain network. The global signal removal for the whole-brain network obtained from the HCP data is the same except that we computed the first global signal from the $\tilde{N} = 116$ ROIs of the AAL atlas (see "Human Connectome Project data" section).

Estimation of discrete states

There are various methods for estimating microstates in the EEG and MEG data [5–7, 13]. We tailor seven popular methods for finding microstates in EEG and MEG data to the case of fMRI data to estimate their discrete states. Because the discrete states that we find for fMRI data are not equivalent to EEG/MEG microstates, we refer to the former as states, discrete states, or clusters in the following text. We describe each method in the following subsections. See Table 1 for main notations.

K-means clustering

The K-means clustering is a simple and popular clustering method to partition the data points into K mutually exclusive clusters. Various EEG and MEG microstate analysis [6, 13, 66] and the studies on temporal variability

Table 1 Main notations used in this paper

Symbols	Description
N_p	Number of participants
N_s	Number of sessions for each participant
N	Number of ROIs
T	Number of volumes (i.e., time points) in each session
$\mathbf{x}_t \in \mathbb{R}^N$	fMRI signal at time t
$\bar{x}_t$	Average of $x_{t,i}$ over ROIs
σ_t	Standard deviation of $x_{t,i}$ over ROIs
K	Number of discrete states
$\mathbf{c}_\ell \in \mathbb{R}^N$	Centroid of the ℓth cluster, where $\ell \in \{1, 2, \ldots, K\}$
L_t	Cluster label for $\mathbf{x}_t$; $L_t \in \{1, 2, \ldots, K\}$

of functional connectivity states of fMRI data [39, 67, 68] used the K-means clustering. It starts with a predefined number of clusters, K. We initialize the centroids of the clusters by the k-means++ algorithm [69]. The k-means++ algorithm consists of the following steps. In step (i), we select one centroid uniformly at random from all the data points. In step (ii), for each data point $\mathbf{x}_t$ that is not yet selected as a centroid, we calculate the distance to the nearest centroid. In step (iii), we sample one $\mathbf{x}_t$ that is not yet selected as a centroid yet with the probability proportional to the square of the distance between $\mathbf{x}_t$ to the nearest centroid. In step (iv), we add the $\mathbf{x}_t$ sampled in the third step as a new centroid. We repeat steps (ii), (iii), and (iv) until we obtain K centroids. This initialization method accelerates the convergence of the algorithm.

Then, we refine the K centroids, denoted by $\mathbf{c}_1, \cdots, \mathbf{c}_K$, as follows. The first step is to assign each data point $\mathbf{x}_t$ to the nearest centroid, i.e., the centroid realizing

$$L_t = \arg\min_{\ell \in \{1, \ldots, K\}} \|\mathbf{x}_t - \mathbf{c}_\ell\|, \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean norm. The second step is to update the centroid of each cluster ℓ by the average of the data points belonging to the cluster as follows:

$$\mathbf{c}_\ell = \frac{\sum_{t=1}^{T} \mathbf{x}_t \delta_{L_t,\ell}}{\sum_{t=1}^{T} \delta_{L_t,\ell}}, \tag{4}$$

where $\delta_{L_t,\ell}$ is the Kronecker delta; $\delta_{L_t,\ell} = 1$ if $L_t = \ell$, and $\delta_{L_t,\ell} = 0$ otherwise. The Kronecker delta in the equation allows us to take the summation only over the data points belonging to the ℓth cluster. We repeat the first and second steps until the change in the residual sum of squares (RSS), defined by

$$\text{RSS} = \sum_{t=1}^{T} \sum_{\ell=1}^{K} \delta_{L_t,\ell} \|\mathbf{x}_t - \mathbf{c}_\ell\|^2, \tag{5}$$

between the two subsequent steps falls below 10^{-5} for the first time. We use the implementation of k-means in scikit-learn [70].

K-medoids clustering
The K-medoids clustering algorithm [71] is a variant of the K-means clustering. The K-medoids clustering uses the original data points as the centroids of the clusters, referred to as medoids. In contrast, the K-means clustering uses the average of the points in the cluster as the centroid of the cluster. The K-medoids clustering begins with a set of K data points as medoids, which we select using the k-medoids++ method. In fact, k-medoids++ is the same as k-means++. In the next step, we assign each $\mathbf{x}_t$ to the ℓth cluster whose medoid is closest to $\mathbf{x}_t$ in terms of the

Euclidean distance. Then, we update the medoid of each cluster to $\mathbf{x}_t$ that belongs to the cluster and minimizes the sum of the Euclidean distance to the other data points in the same cluster. We repeat the last two steps until the dissimilarity score (i.e., the sum of the Euclidean distance from the medoid to the other data points in the cluster) stops changing for each cluster. We use the k-medoids implemented in scikit-learn [70].

Agglomerative hierarchical clustering
Agglomerative hierarchical clustering, which we simply call the agglomerative clustering (AC), is a bottom-up clustering method. The AC method initially regards each data point as a single-node cluster. Then, one merges a pair of clusters one after another based on a linkage criterion. Among various linkage criteria, we use the Ward's method implemented in scikit-learn [70]. In each step of merging two clusters, the Ward's method minimizes the within-cluster variance, i.e., the squared Euclidean distance between $\mathbf{x}_t$ and the centroid of the new cluster to which $\mathbf{x}_t$ belongs, which is summed over $t \in \{1, \ldots, T\}$. We stop the cluster merging procedure once the number of clusters is equal to K.

Atomize and agglomerate hierarchical clustering
The Atomize and Agglomerate Hierarchical Clustering (AAHC) is another bottom-up hierarchical clustering algorithm [9, 72, 73]. A main difference between AAHC and traditional bottom-up hierarchical clustering methods is that AAHC atomizes the worst cluster. In other words, AAHC disintegrates the worst cluster and assigns each member of this cluster to a different cluster instead of merging the entire worst cluster with the most similar cluster.

AAHC uses the global explained variance (GEV) as a measure of the quality of the cluster [6, 9, 73, 74]. The GEV for the ℓth cluster is defined by

$$\text{GEV}_\ell = \frac{\sum_{t=1}^{T} \delta_{L_t,\ell} \, \text{corr}(\mathbf{x}_t, \mathbf{c}_\ell)^2 \, \sigma_t^2}{\sum_{t=1}^{T} \sigma_t^2}, \tag{6}$$

where $\text{corr}(\mathbf{x}_t, \mathbf{c}_\ell)$ is the cosine similarity between $\mathbf{x}_t$ and $\mathbf{c}_t$ given by

$$\text{corr}(\mathbf{x}_t, \mathbf{c}_\ell) = \frac{\langle \mathbf{x}_t, \mathbf{c}_\ell \rangle}{\|\mathbf{x}_t\| \|\mathbf{c}_\ell\|}. \tag{7}$$

In Eq. (7), $\langle \mathbf{x}_t, \mathbf{c}_\ell \rangle$ is the inner product of $\mathbf{x}_t$ and $\mathbf{c}_\ell$. Variable σ_t represents the standard deviation of the data point $\mathbf{x}_t$ across the ROIs and is given by Eq. (2). Quantity σ_t is known as global field power (GFP) in the literature of microstate analysis for EEG and MEG data [5, 13, 73, 75]. For the second and third treatments of the global signal removal, it holds true that $\sigma_t = 1$ for any t because of the

Islam *et al. BMC Neuroscience*   (2024) 25:14

Page 6 of 19

global signal removal carried out in the last step of the treatment.

In the AAHC, we define the worst cluster as the one with the smallest GEV_ℓ, $\ell \in \{1, 2, \cdots, K\}$ and atomize it. Then, we assign each data point $\mathbf{x}_t$ of the atomized cluster to the ℓth cluster that maximizes Eq. (7) [9, 73]. As in the AC, the AAHC initially regards each $\mathbf{x}_t$ as a single-node cluster. We repeat finding the worst cluster, atomizing it, and assigning each $\mathbf{x}_t$ in the atomized cluster to a different cluster until the number of clusters reaches K.

Topographic atomize and agglomerate hierarchical clustering

The Topographic Atomize and Agglomerate Hierarchical Clustering (TAAHC) is a modification of AAHC [9, 74]. The difference between AAHC and TAAHC is that TAAHC defines the worst cluster to be the ℓth cluster that is the smallest in terms of the sum of the correlation of the data points in the cluster with its centroid $\mathbf{c}_\ell$ [9, 74]. In other words, the worst cluster ℓ is the minimizer of

$$\text{CRS}(\ell) = \sum_{t=1}^{T} \delta_{L_t, \ell} \, \text{corr}(\mathbf{x}_t, \mathbf{c}_\ell) = \sum_{t=1}^{T} \frac{\delta_{L_t, \ell} \langle \mathbf{x}_t, \mathbf{c}_\ell \rangle}{\|\mathbf{x}_t\| \|\mathbf{c}_\ell\|} \quad (8)$$

over $\ell \in \{1, \ldots, K\}$. As in the AC and AAHC, the TAAHC first regards each $\mathbf{x}_t$ as a single-node cluster. Second, we identify the cluster with the smallest $\text{CRS}(\ell)$. Third, we atomize the selected cluster and reassign each of its member $\mathbf{x}_t$ to the cluster whose centroid is the closest to $\mathbf{x}_t$ in terms of $\text{corr}(\mathbf{x}_t, \mathbf{c}_\ell)$. We iterate the second and third steps until we obtain K clusters.

Bisecting K-means clustering

The bisecting K-means method combines the K-means clustering method and divisive hierarchical clustering [76]. Initially, we let all data points form a single cluster. Then, we apply the K-means clustering with $K = 2$ to partition the data points into two clusters, by following the procedure described in "K-means clustering" section. Then, we select the cluster that has the larger value of the dissimilarity defined for the ℓth cluster by

$$\text{SSE}_\ell = \sum_{t=1}^{T} \delta_{L_t, \ell} \|\mathbf{x}_t - \mathbf{c}_\ell\|^2. \quad (9)$$

Then, we run the K-means clustering on the selected cluster to split it into two clusters. We repeat selecting the cluster with the largest SSE_ℓ and bisecting it until we obtain K clusters. We use the implementation of the bisecting K-means in scikit-learn [70].

Gaussian mixture model

The Gaussian mixture model (GMM) represents each cluster as a multivariate Gaussian distribution. We denote by $\mathcal{N}(\mu_\ell, \Sigma_\ell)$, with $\ell \in \{1, 2, \ldots, K\}$, the multidimensional Gaussian distribution with mean vector μ_ℓ and covariance matrix Σ_ℓ [22, 77]. The GMM is given by

$$p(\mathbf{x}_t) = \sum_{\ell=1}^{K} \pi_\ell \, \mathcal{N}(\mathbf{x}_t | \mu_\ell, \Sigma_\ell), \quad (10)$$

where π_ℓ is the mixing weight, i.e., the probability that a data point originates from the ℓth multivariate Gaussian distribution. Note that $\sum_{\ell=1}^{K} \pi_\ell = 1$. The likelihood function for the set of all the data points is given by

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_T) = \prod_{t=1}^{T} \sum_{\ell=1}^{K} \pi_\ell \, \mathcal{N}(\mathbf{x}_t | \mu_\ell, \Sigma_\ell). \quad (11)$$

We infer the parameter values by maximizing the log-likelihood function using an expectation-maximization (EM) algorithm [22, 77, 78]. We regard μ_ℓ as the centroid of the ℓth cluster. Because the GMM is a soft clustering method, we assign each time point t to the ℓth cluster that maximizes $\hat{\pi}_\ell \mathcal{N}(\mathbf{x}_t | \hat{\mu}_\ell, \hat{\Sigma}_\ell)$, where $\hat{\pi}_\ell$, $\hat{\mu}_\ell$, and $\hat{\Sigma}_\ell$ are the obtained maximum likelihood estimator. We use the GaussianMixture class in scikit-learn, which uses K-means clustering for initializing the parameters [70].

Among the seven methods that we employ to cluster the fMRI data, the GMM is the only parametric model. All the other methods are non-parametric clustering methods.

Evaluation of the clustering methods

The number of microstates estimated for EEG and MEG data depends on studies [6, 9, 66, 75]. Studies on temporal dynamics of functional connectivity in fMRI data are also diverse with respect to the number of clusters [39, 67, 68]. Therefore, we examine the number of states, K, from 2 to 10 for each clustering algorithm. To compare the quality of the different clustering methods, we use the GEV given by Eq. (6). The GEV captures the amount of the data variance explained by the microstates' centroids, also called the global map, cluster map, microstate map, and template map [7, 9, 13, 73]. We calculate the total GEV as the sum of the GEV over all the states, i.e.,

$$\text{GEV}_{\text{total}} = \sum_{\ell=1}^{K} \text{GEV}_\ell \quad (12)$$

and average it over all the sessions and participants. A large value of the $\text{GEV}_{\text{total}}$ suggests that the obtained clustering is of high quality.

Islam *et al. BMC Neuroscience*   (2024) 25:14

Page 7 of 19

We also measure the quality of the clustering methods using the within-cluster sum of squares (WCSS) [79], also known as the distortion measure [77]. The WCSS is defined by

$$\text{WCSS} = \sum_{t=1}^{T} \sum_{\ell=1}^{K} \delta_{L_t,\ell} \|\mathbf{x}_t - \mathbf{c}_\ell\|^2. \tag{13}$$

A small WCSS value indicates that the data points are tightly clustered and therefore the clustering is of high quality.

Comparison of state-transition dynamics between different sessions

Observables for the state-transition dynamics

To test reproducibility of the fMRI state-transition dynamics across participants and sessions, we measure the following five observables for each session. These observables are often used in the analysis of microstate dynamics for EEG and MEG data [9, 13, 14, 66] and activity patterns for fMRI data [32, 39, 67].

First, we use the centroid of each of the K states as an observable. The centroid $\mathbf{c}_\ell$ of the ℓth state represents the set of data points which are assigned to the ℓth state. We remind that the centroid is an N-dimensional vector.

Second, the coverage time of the ℓth state is the number of times $t \in \{1, \ldots, T\}$ in which the ℓth state appears. We normalize the coverage time of each state by dividing it by the total observation time, T.

Third, we measure the frequency of appearance of each state. If the ℓth state starts and then lasts for some time steps before transiting to a different state, then we say that this is a unique appearance of ℓ. That is, we count the consecutive appearances as one unique appearance. The frequency of appearance of ℓ is defined as the number of unique appearance divided by T.

Fourth, the average lifespan of the ℓth state is the time spent in a unique appearance of ℓ that is averaged over all unique appearances of ℓ. The average lifespan of ℓ is equal to the coverage time divided by the number of unique appearance of ℓ.

Fifth, we investigate the frequency of transitions from one state to another as follows. Let $n_{\ell\ell'}$ be the number of times with which the transition from the ℓth state to the ℓ'th state occurs in the given session, where $\ell' \neq \ell$. We define the transition probability from ℓ to ℓ' by $p_{\ell\ell'} = n_{\ell\ell'} / \sum_{\ell''=1; \ell'' \neq \ell}^{K} n_{\ell\ell''}$ and we set $p_{\ell\ell} = 0$. The $K \times K$ transition probability matrix is given by $P = (p_{\ell\ell'})$ with $\ell, \ell' \in \{1, \ldots, K\}$.

Discrepancy measures for comparing the state-transition dynamics between two sessions

For examining the reproducibility of state-transition dynamics between sessions of the same participant and between different participants, we need to compare observables between pairs of sessions. To this end, we first need to find the best matching of the states between the two sessions. For $K \in \{2, \ldots, 8\}$, we assess all the possible pairwise matchings of the states between the two sessions. This entails exhaustively matching every permutation of the K states of the one session with the states of the another session. The total number of such pairwise matchings is equal to $K!$. For each matching, we calculate the correlation between centroids $\mathbf{c}_\ell$ and $\mathbf{c}_{\ell'}$ of the matched states, i.e., ℓth state in the first session and the ℓ'th state in the second session, by corr$(\mathbf{c}_\ell, \mathbf{c}_{\ell'})$, where corr is defined in Eq. (7). We then average corr$(\mathbf{c}_\ell, \mathbf{c}_{\ell'})$ over all the K matched pairs of states in the two sessions and call it the centroid similarity. We select the matching that maximizes the centroid similarity among the $K!$ matchings.

For $K = 9$ and $K = 10$, we cannot assess all possible pairwise matchings due to combinatorial explosion. Therefore, we use a greedy search to find an approximately optimal matching. First, we find the pair of the ℓth state in the first session and the ℓ'th state in the second session that maximizes corr$(\mathbf{c}_\ell, \mathbf{c}_{\ell'})$. Second, we select one state from the remaining $K - 1$ states in the first session and one state from the remaining $K - 1$ states in the second session such that the correlation between the two centroids is the largest, and we pair them. We repeat this procedure until all the K states are matched between the two sessions.

Once we have determined the final matching between the K states in the first session and those in the second session, we use the centroid dissimilarity, defined as $1 - (\text{centroid similarity})$, as a measure of discrepancy between the set of K states in the two sessions. The centroid dissimilarity ranges between 0 and 2. It is equal to 0 if and only if the set of the L centroid positions is exactly parallel between the two sessions.

The centroid similarity, corr$(\mathbf{c}_\ell, \mathbf{c}_{\ell'})$, only compares the direction of the two centroids, $\mathbf{c}_\ell$ and $\mathbf{c}_{\ell'}$, from the origin. Therefore, we also measured the discrepancy between the set of K states in the two sessions based on the Euclidean distance between $\mathbf{c}_\ell$ and $\mathbf{c}_{\ell'}$, given by

$$d(\mathbf{c}_\ell, \mathbf{c}_{\ell'}) = \|\mathbf{c}_\ell - \mathbf{c}_{\ell'}\|^2. \tag{14}$$

In the verification analysis, we searched for the best matching of the K states between the two sessions by minimizing the average of $d(\mathbf{c}_\ell, \mathbf{c}_{\ell'})$ over the K matched pairs of states instead of maximizing the average of corr$(\mathbf{c}_\ell, \mathbf{c}_{\ell'})$. Similar to the case of using corr$(\mathbf{c}_\ell, \mathbf{c}_{\ell'})$, we did so by the exhaustive search when $K \in \{2, \ldots, 8\}$

and by the greedy algorithm when $K \in \{9, 10\}$. The dissimilarity obtained using the average of d is equal to 0 if and only if the set of the L centroid positions is the same between the two sessions, and its large value implies a large discrepancy between the two sessions in terms of the centroid position.

For the coverage time, frequency of appearance, and average lifespan of states, we compute the total variation (TV) to quantify the difference in the state-transition dynamics between two sessions. Let $Q_i(\ell)$ be the coverage time, frequency of appearance, or average lifespan for the ℓth state in session i. For the notational convenience, we assume without loss of generality that we have matched the ℓth state in session i with the ℓth state in session j. For the coverage time of the ℓth microstate, we use the normalized coverage time defined in "Observables for the state-transition dynamics" section as $Q_i(\ell)$. The TV is defined by

$$\delta(Q_i, Q_j) = \max_{\ell \in \{1,2,\ldots,K\}} \left| Q_i(\ell) - Q_j(\ell) \right|, \tag{15}$$

where $Q_i = \{Q_i(1), \ldots, Q_i(K)\}$.

To quantify the difference between the transition probability matrices for two sessions i and j, denoted by $P^{(i)} = \left(p_{\ell\ell'}^{(i)} \right)$ and $P^{(j)} = \left(p_{\ell\ell'}^{(j)} \right)$, respectively, where $\ell, \ell' \in \{1, \ldots, K\}$, we calculate the Frobenius distance given by

$$\left\| P^{(i)} - P^{(j)} \right\|_{\mathrm{F}} = \sqrt{\sum_{\ell=1}^{K} \sum_{\ell'=1}^{K} \left| P_{\ell\ell'}^{(i)} - P_{\ell\ell'}^{(j)} \right|^2}. \tag{16}$$

Permutation test

We hypothesize that the state-transition dynamics estimated from fMRI data is more consistent between different sessions of the same participant than between different participants. To test this hypothesis, we compare the dissimilarity between two sessions originating from the same participant and the dissimilarity between two sessions originating from different participants. If the former is smaller than the latter, then the state-transition dynamics is more reproducible within a participant than between different participants, supporting the potential ability of state-transition dynamics to be used for individual fingerprinting.

We measure the dissimilarity between a given pair of sessions in terms of one of the five observables (i.e., centroid position, distribution of the coverage time, normalized frequency of appearance of states, distribution of the average lifespan, or the transition probability matrix). For each observable, we compare the within-participant dissimilarity and between-participant dissimilarity using the

normalized distance ND combined with the permutation test [80, 81], which we adapt here for our purpose. Denote by $q(p, s)$ one of the five observables for participant $p \in \{1, \ldots, N_{\mathrm{p}}\}$ and session $s \in \{1, \ldots, N_{\mathrm{s}}\}$, where $N_{\mathrm{p}} = 8$ is the number of participants, and $N_{\mathrm{s}} = 10$ is the number of sessions per participant. We define the ND by

$$\mathrm{ND}(q) = \frac{\frac{2}{N_{\mathrm{p}}(N_{\mathrm{p}}-1)N_{\mathrm{s}}} \sum_{s=1}^{N_{\mathrm{s}}} \sum_{p=1}^{N_{\mathrm{p}}} \sum_{p'=1}^{p-1} \tilde{d}\left(q(p,s), q(p',s)\right)}{\frac{2}{N_{\mathrm{p}}N_{\mathrm{s}}(N_{\mathrm{s}}-1)} \sum_{p=1}^{N_{\mathrm{p}}} \sum_{s=1}^{N_{\mathrm{s}}} \sum_{s'=1}^{s-1} \tilde{d}\left(q(p,s), q(p,s')\right)}$$

$$= \frac{(N_{\mathrm{s}}-1) \sum_{s=1}^{N_{\mathrm{s}}} \sum_{p=1}^{N_{\mathrm{p}}} \sum_{p'=1}^{p-1} \tilde{d}\left(q(p,s), q(p',s)\right)}{(N_{\mathrm{p}}-1) \sum_{p=1}^{N_{\mathrm{p}}} \sum_{s=1}^{N_{\mathrm{s}}} \sum_{s'=1}^{s-1} \tilde{d}\left(q(p,s), q(p,s')\right)}, \tag{17}$$

where $\tilde{d}$ denotes the dissimilarity (i.e., the Euclidean distance, TV, or Frobenius norm, depending on the observable; see "Discrepancy measures for comparing the state-transition dynamics between two sessions" section) between two sessions. The prefactor on the right-hand side on the first line of Eq. (17) accounts for the normalization; there are $\frac{N_{\mathrm{p}}(N_{\mathrm{p}}-1)N_{\mathrm{s}}}{2}$ and $\frac{N_{\mathrm{p}}N_{\mathrm{s}}(N_{\mathrm{s}}-1)}{2}$ terms in the summation on the numerator and denominator, respectively. Therefore, the numerator of the right-hand side on the first line of Eq. (17) represents the average dissimilarity between two sessions obtained from different participants. The denominator represents the average dissimilarity between two sessions obtained from the same participant. If the state-transition dynamics are more consistent among different sessions within the same participant than among different sessions of different participants, we expect that $\mathrm{ND}(q) > 1$.

To statistically test the $\mathrm{ND}(q)$ value, we ran a permutation test [82]. Specifically, we carried out the following steps.

(i) Shuffle the values of q across all participants and sessions uniformly at random. This process is equivalent to applying a random permutation on $\{q(1, 1), q(1, 2), \ldots, q(N_{\mathrm{p}}, N_{\mathrm{s}})\}$. We denote the q value for the sth session for pth participant after the random permutation by $q'(p, s)$. Note that the q' value originates from any of the N_{p} participants with probability $1/N_{\mathrm{p}}$ and any of the N_{s} sessions with probability $1/N_{\mathrm{s}}$.

(ii) Calculate $\mathrm{ND}(q')$.

(iii) Repeat steps (i) and (ii) R times. We set $R = 10^4$.

(iv) The permutation p-value is equal to the fraction of the runs among the R runs in which the $\mathrm{ND}(q')$ value is larger than the empirical $\mathrm{ND}(q)$ value.

Results

Choice of the global signal reduction and clustering methods

We ran the seven clustering methods for each number of clusters, $K \in \{2, \ldots, 10\}$, each of the ten sessions, each of the eight participants, and each of the three global signal removal methods for the DMN extracted from the MSC data. Then, we calculated the total GEV, i.e., $\text{GEV}_{\text{total}}$, for each combination of these variables as a measure of the quality of clustering. We show the $\text{GEV}_{\text{total}}$ values averaged over all the participants and sessions in Fig. 1a–c, for each combination of these variations. Each panel of Fig. 1 corresponds to a treatment of the global signal removal. In all the cases, $\text{GEV}_{\text{total}}$ increases as K increases. We also find that $\text{GEV}_{\text{total}}$ is notably larger with the first and second treatments than the third treatment for all the seven clustering methods and that $\text{GEV}_{\text{total}}$ is slightly larger under the second than the first treatment for all values of K and for all the clustering methods. Because the second treatment of the global signal removal shows the best performance in terms of clustering quality (i.e., providing the largest $\text{GEV}_{\text{total}}$), we use the second treatment in the following analyses.

We select the three clustering methods with the largest $\text{GEV}_{\text{total}}$, which are the K-means, TAAHC, and bisecting

K-means. For these three clustering methods, $\text{GEV}_{\text{total}}$ is around 70% with $K = 4$ (K-means: $70.22 \pm 2.76\%$ (average $\pm$ standard deviation calculated on the basis of the 80 sessions of the MSC data), TAAHC: $68.56 \pm 2.98\%$, bisecting K-means: $69.48 \pm 2.93\%$) and more than 75% with $K = 7$ (K-means: $77.49 \pm 2.07\%$, TAAHC: $76.16 \pm 2.29\%$, bisecting K-means: $75.85 \pm 2.31\%$). We show a brain map of each of the $K = 4$ discrete states estimated by K-means in Fig. 2. As references, previous microstate analyses on EEG data found the $\text{GEV}_{\text{total}}$ values of $70.92 \pm 3.65\%$ [9] and $65.80 \pm 4.90\%$ [7] using the K-means clustering, and $69.93 \pm 3.58\%$ using TAAHC [9], all with $K = 4$. Furthermore, $\text{GEV}_{\text{total}}$ values of $65.03 \pm 6.13\%$ and $60.99 \pm 5.62\%$ with $K = 5$ were reported for EEG data recorded under eyes-closed and eyes-open conditions, respectively [66]. A MEG study reported a $\text{GEV}_{\text{total}}$ of $63.97 \pm 0.64\%$ using the K-means clustering with $K = 10$ [13]. Our present data analysis with the fMRI data has yielded somewhat larger $\text{GEV}_{\text{total}}$ values than these studies.

The GEV is based on the similarity in the direction of the N-dimensional fMRI signal, $\mathbf{x}_t$, and the centroid of the cluster, $\mathbf{c}_{L_t}$, where we remind that L_t is the index of the cluster to which $\mathbf{x}_t$ belongs. Therefore, the GEV can be large even if $\mathbf{x}_t$ and $\mathbf{c}_{L_t}$ are not close to each other.
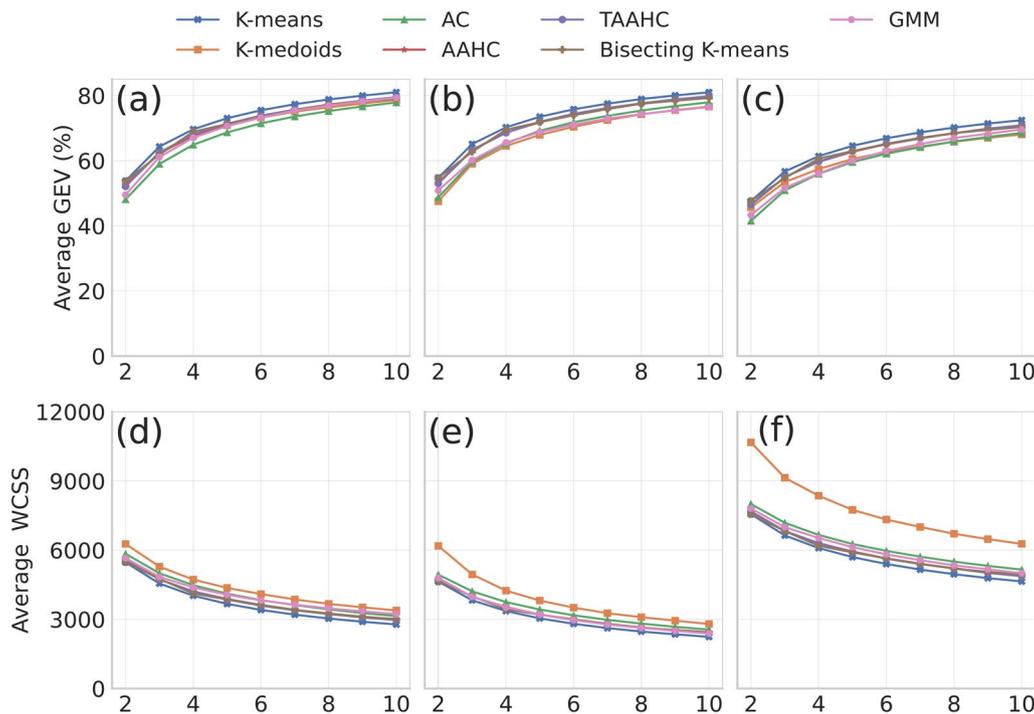
Fig. 1 Performance of estimating discrete states from the DMN extracted from the MSC data. We show the results for the three treatments of global signal removal, seven clustering methods, and $K \in \{2, \ldots, 10\}$. **a**–**c** Total GEV. **d**–**f** WCSS. **a**, **d** First treatment of the global signal removal. **b**, **e** Second treatment. **c**, **f** Third treatment. Each $\text{GEV}_{\text{total}}$ and WCSS value shown is the average over the eight participants and ten sessions per participant
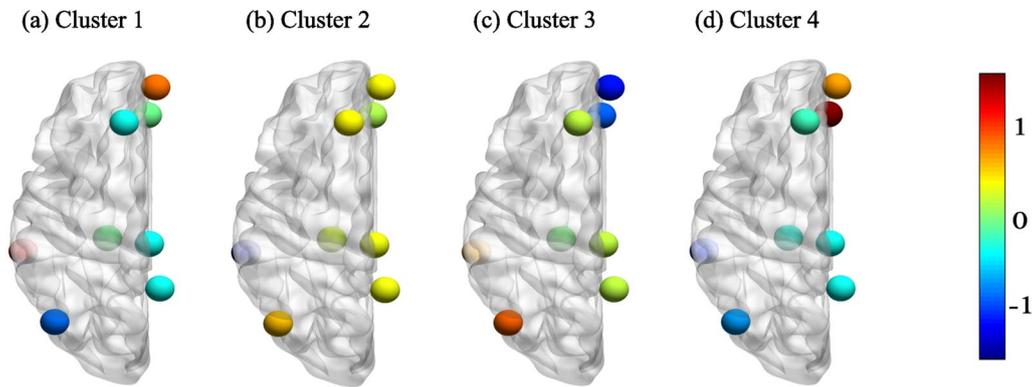
Fig. 2 Brain map for each of the $K=4$ discrete states estimated by K-means for the first session of the first participant (i.e., MSC01). This choice of session and participant is arbitrary. We employed the second treatment of the global signal removal. Each circle represents an ROI in the left hemisphere. The color represents the activity averaged over the volumes belonging to the corresponding state. We used the BrainNet Viewer tool [85] for the visualization

Therefore, we also computed the WCSS, which is the sum of the distance between $\mathbf{x}_t$ and $\mathbf{c}_{L_t}$ over all the volumes. We confirmed that the dependence of the WCSS on the global signal removal method, clustering method, and K is similar to that with GEV_{total} (see Fig. 1d–f). Note that a large GEV_{total} value implies a good clustering result, whereas a small WCSS value implies a good clustering result. In particular, with the WCSS, the second treatment of the global signal removal is the best among the three treatments, and the best three clustering methods remain the same, while the GMM performs as equally well as the TAAHC and the bisecting K-means for the second treatment of the global signal removal (see Fig. 1e). Therefore, in the remainder of this paper, we further focus our analysis only on the K-means, TAAHC, and bisecting K-means clustering methods.

Global signals can show spatio-temporal patterns and may provide specific pathological or psychological information [83, 84]. Therefore, we examined the quality of clustering produced by seven clustering methods and for $K \in \{2, \cdots, 10\}$ in the absence of global signal removal. The quality of clustering was worse without global signal removal than with global signal removal in terms of both GEV_{total} and WCSS (see Additional file 1: section S2). Therefore, we do not consider the analysis without global signal removal in the following sections.

Test–retest reliability of the observables of the state-transition dynamics
We calculated the five observables, i.e., centroid of the clusters, coverage time, frequency, average lifespan, and transition probability matrix, of the estimated state-transition dynamics for each of the three selected clustering methods, each K value, session, and participant. Then, we calculated the discrepancy in each observable between two sessions. To compare the state-transition dynamics of different sessions within the same participant, which we call the within-participant comparison, we calculated the discrepancy in terms of each observable for each pair of sessions for each participant. Because there are ten sessions for each of the eight participants, there are $\binom{10}{2} \times 8 = 360$ within-participant comparisons, where $\binom{}{}$ represents the binomial coefficient. To compare the state-transition dynamics between different participants, which we call the between-participant comparison, we calculated the discrepancy in terms of each observable between each pair of sessions obtained from different participants. There are $\binom{8}{2} \times 10 = 280$ between-participant comparisons.

We show the distribution of the discrepancy measure for each observable with $K = 4$, separately for the within-participant and between-participant comparisons, in Fig. 3. Figure 3a–c shows the results for the K-means, TAAHC, and bisecting K-means, respectively. We find that the state-transition dynamics are visually more similar in the within-participant than between-participant comparisons across all the indices and for all the three clustering methods when we compare the minimum, maximum, median, first quartile, and third quartile values of each distribution. For all the three clustering methods, the gap between the within-participant and between-participant comparison is apparently the largest for the centroid position among the five observables. The gap between the within-participant and between-participant comparisons often looks subtle, in particular for the coverage time. The results with $K = 7$ and $K = 10$ are qualitatively the same as those with $K = 4$ (see Additional file 1: section S2).
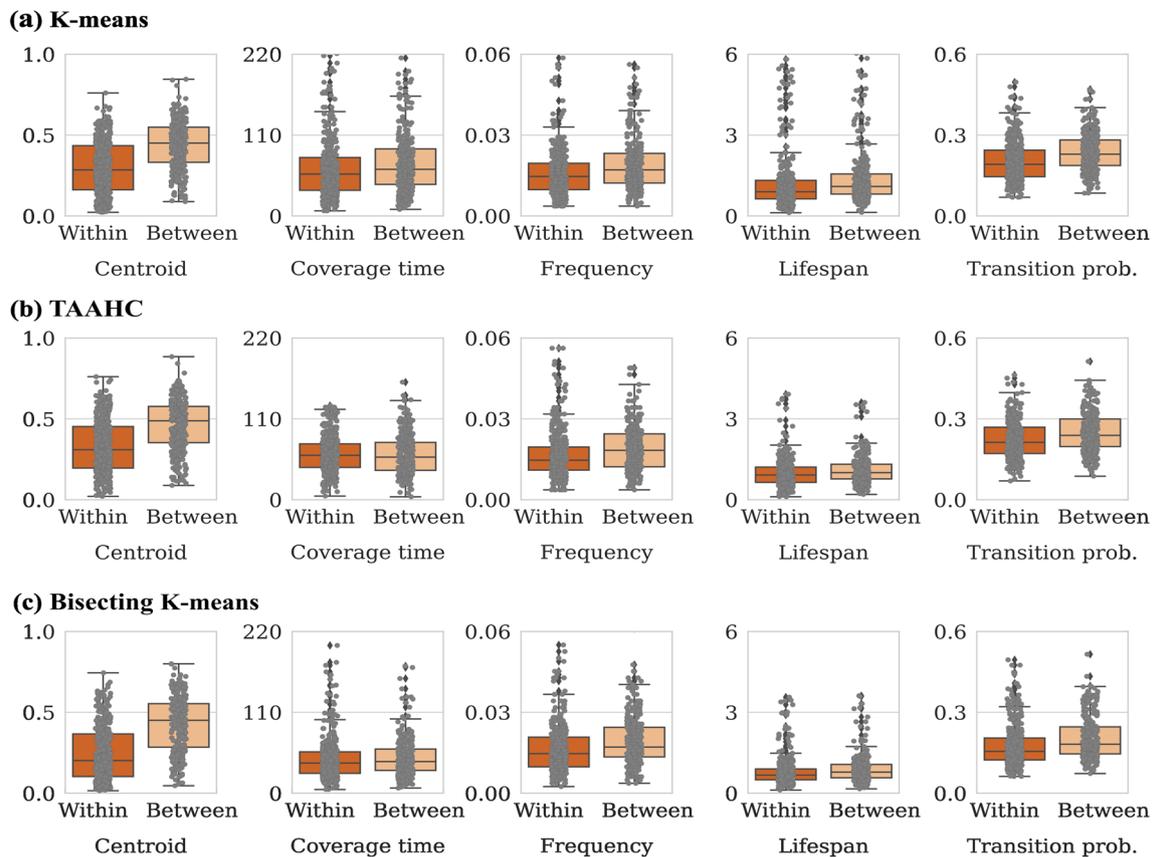
Fig. 3 Within-participant and between-participant reproducibility of the state-transition dynamics with $K = 4$ states. **a** K-means. **b** TAAHC. **c** Bisecting K-means. "Within" and "Between" indicate the within-participant and between-participant comparisons, respectively. Each box plot shows the minimum, maximum, median, first quartile, and third quartile of the measurements. Each dot represents a session. "Centroid" abbreviates the centroid position, and "Transition prob." abbreviates the transition probability matrix

To test the significance of the difference between the within-participant and between-participant session-to-session reproducibility of the state-transition dynamics, we carried out the permutation test. We computed the ND value for each clustering method, value of $K \in \{2, \ldots, 10\}$, and observable. Furthermore, we computed the ND values for 10^4 randomized session-to-session comparisons. The permutation test concerns whether the ND value for the original session-to-session comparisons is significantly different from the ND values for the comparisons between the randomized pairs of sessions.

With $K = 4$, we show the ND value for the original session-to-session comparisons and the distribution of the ND values for the randomized sessions in Fig. 4. Each panel of Fig. 4 shows a combination of the clustering method and the observable. The vertical dashed lines represent the ND values for the original session-to-session comparisons. We find that the result of the permutation test is significant in many cases even after correcting for multiple comparisons over the three clustering methods,

nine values of K, and five observables; see the uncorrected p values in the figure; an uncorrected $p = 0.00037$ corresponds to a Bonferroni corrected $p = 0.05$. A small p value implies that the within-participant session-to-session reproducibility is higher than the between-participant session-to-session reproducibility, suggesting the possibility of using the observable for fingerprinting individuals.

We tabulate the p values from the permutation test for the three clustering methods, $K \in \{2, \ldots, 10\}$, and the five observables in Table 2. In the table, $p < 10^{-4}$ indicates that the ND value for the original session-to-session comparisons is farther from 1 than all the 10^4 randomized comparisons. The table shows that a majority of the p values (i.e., 126 out of 135; 93.33%) are smaller than 0.05 (shown with *). One hundred and seventeen of them (i.e., 86.67% of the 135 comparisons) remain significant after the Bonferroni correction (shown with ***; equivalent to $p < 0.00037$, uncorrected). Because there are 135 comparisons in the table and using the Bonferroni correction may be too stringent, we also counted the
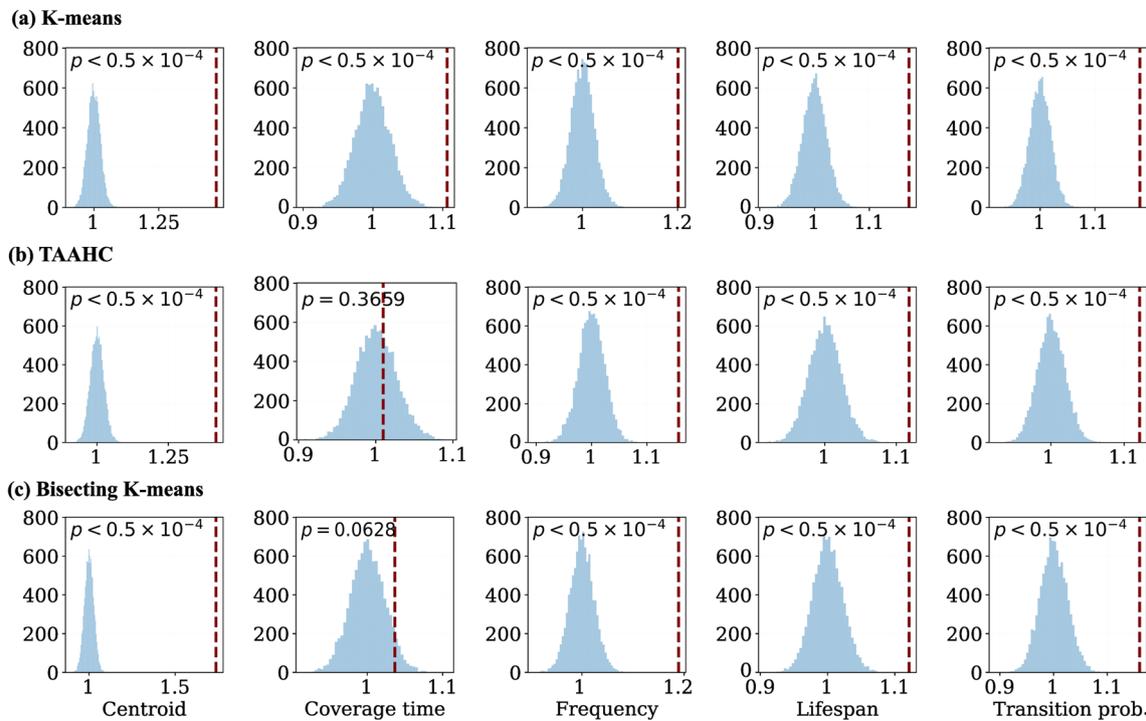
Fig. 4 Distribution of the ND values for the original and randomized session-to-session comparisons with $K = 4$. **a** K-means. **b** TAAHC. **c** Bisecting K-means. The p values shown are the uncorrected values

cases in which the uncorrected p value is less than 0.001, shown with **; there are 119 out of the 135 comparisons (i.e., 88.15%) with $p < 0.001$. We find that the number of significant p values with the K-means and bisecting K-means is somewhat larger than with the TAAHC (with $p < 0.05$, K-means, TAAHC, and bisecting K-means have 44, 39, and 43 significant comparisons, respectively). We also find that the p values considerably depend on the observables. The permutation test result is strongly significant (i.e., $p < 10^{-4}$) for all the clustering methods and K values for the centroid position. In contrast, the number of significant combinations of the clustering method and K value is smallest for the coverage. Lastly, we do not observe a notable dependence of the permutation test result on K.

Robustness tests

For validation, we also estimated the state-transition dynamics for the whole-brain network extracted from the MSC data. We show the results for the permutation test in Table 3. The results are similar to those for the DMN. In particular, the centroid position is the most effective among the five observables at distinguishing between the within-participant and between-participant comparisons, and the coverage is the least effective.

To examine the robustness of the proposed method with respect to fMRI experiments, we also ran the same permutation test for a whole-brain network obtained from the resting-state HCP data. Note that, among various parameters, the TR for the HCP data (i.e., 0.72 s) is substantially different from that for the MSC data (i.e., 2.2 s). It should also be noted that we use different atlases for the MSC and HCP data. While this decision is convenient for us because we have been using the obtained ROI-based fMRI data in different projects, it allows us to investigate the robustness of the proposed methods with respect to the atlas. The results, shown in Table 4, are similar to those for the DMN and whole-brain networks extracted from the MSC data. However, the permutation test results were stronger for the HCP than MSC data (i.e., a larger number of significant comparisons among the 135 comparisons). In particular, the frequency, lifespan, and the transition probability matrix as well as the centroid position yielded the smallest possible p value (i.e., $p < 10^{-4}$) for all pairs of the clustering method and K value.

Moreover, as we noted earlier, our main definition of the centroid dissimilarity relies on the (dis)similarity between $\mathbf{c}_{L_t}$ and $\mathbf{x}_t$ only in terms of the direction. Therefore, we reran the permutation test by replacing the centroid (dis)similarity by the WCSS to measure the

Table 2 Results of the permutation test for the DMN extracted from the MSC data

	K	Centroid	Coverage	Frequency	Lifespan	Trans. prob.
K-means	2	$< 10^{-4}$***	0.1975	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	3	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	4	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	5	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	6	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	7	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	8	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	9	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	10	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
TAAHC	2	$< 10^{-4}$***	0.1077	0.0001***	$< 10^{-4}$***	0.0001***
	3	$< 10^{-4}$***	0.0001***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	4	$< 10^{-4}$***	0.3659	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	5	$< 10^{-4}$***	0.0928	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	6	$< 10^{-4}$***	0.0162*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	7	$< 10^{-4}$***	0.5488	$< 10^{-4}$***	0.0003***	$< 10^{-4}$***
	8	$< 10^{-4}$***	0.0642	$< 10^{-4}$***	0.0001***	$< 10^{-4}$***
	9	$< 10^{-4}$***	0.0942	0.0001***	$< 10^{-4}$***	$< 10^{-4}$***
	10	$< 10^{-4}$***	0.0132*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
Bisecting K-means	2	$< 10^{-4}$***	0.0568	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	3	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	4	$< 10^{-4}$***	0.0628	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	5	$< 10^{-4}$***	0.0114*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	6	$< 10^{-4}$***	0.0102*	$< 10^{-4}$***	0.0003***	$< 10^{-4}$***
	7	$< 10^{-4}$***	0.0003***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	8	$< 10^{-4}$***	0.0017*	0.0004**	$< 10^{-4}$***	$< 10^{-4}$***
	9	$< 10^{-4}$***	0.0112*	0.0009**	$< 10^{-4}$***	$< 10^{-4}$***
	10	$< 10^{-4}$***	0.0162*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***

*p < 0.05, uncorrected

**p < 0.001, uncorrected

***p < 0.00037, uncorrected (which is equivalent to *p*<0.05, Bonferroni corrected). We remark that "Centroid" and "Trans. prob." abbreviate the centroid's position and the transition probability matrix, respectively

average distance between $\mathbf{c}_{L_t}$ and $\mathbf{x}_t$. This change not only affected the discrepancy measure between two sessions in terms of the centroid position but also the discrepancy between pairs of sessions in terms of the other four observables (i.e., coverage time, frequency of appearance of each state, average lifespan, and transition probability matrix). This is because changing the discrepancy measure for cluster centroids affects how the set of centroids (and therefore clusters) is matched between two given sessions. We confirmed that the permutation test results with the WCSS is similar to those with $\mathrm{GEV}_{\mathrm{total}}$ (see Additional file 1: section S4). In particular, the *p* values were overall small, the results tended to be more significant for the K-means and bisecting K-means than for the

TAAHC, and for the centroid position and the transition probability matrix than for the other three observables.

Lastly, we employed a 59-ROI DMN [52] extracted from MSC data to further assess the impact of the dimension reduction. We found that results of the test–retest reproducibility are roughly as good as those with the 12-ROI DMN (see Additional file 1: section S5).

Discussion

We carried out a comparative study of methods to cluster volumes of the fMRI to extract time series of the system's state, akin to microstate analysis for EEG and MEG data, for each recording session. We found that aggregating the symmetrically located ROIs into one ROI

Islam *et al. BMC Neuroscience*   (2024) 25:14

Page 14 of 19

Table 3 Results of the permutation test for the whole-brain network extracted from the MSC data

	K	Centroid	Coverage	Frequency	Lifespan	Trans. prob.
K-means	2	$< 10^{-4}$***	0.045*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	3	$< 10^{-4}$***	0.0193*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	4	$< 10^{-4}$***	0.0414*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	5	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	6	$< 10^{-4}$***	0.0002***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	7	$< 10^{-4}$***	0.0008**	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	8	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	9	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	10	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
TAAHC	2	$< 10^{-4}$***	0.1747	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	3	$< 10^{-4}$***	0.0167*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	4	$< 10^{-4}$***	0.4553	$< 10^{-4}$***	0.0001***	$< 10^{-4}$***
	5	$< 10^{-4}$***	0.0288*	0.0001***	$< 10^{-4}$***	$< 10^{-4}$***
	6	$< 10^{-4}$***	0.5063	0.0002***	$< 10^{-4}$***	$< 10^{-4}$***
	7	$< 10^{-4}$***	0.7551	0.0008**	$< 10^{-4}$***	$< 10^{-4}$***
	8	$< 10^{-4}$***	0.1887	0.0001***	$< 10^{-4}$***	$< 10^{-4}$***
	9	$< 10^{-4}$***	0.0874	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	10	$< 10^{-4}$***	0.0426*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
Bisecting K-means	2	$< 10^{-4}$***	0.0283*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	3	$< 10^{-4}$***	0.0597	$< 10^{-4}$***	0.0007**	0.0012*
	4	$< 10^{-4}$***	0.0402*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	5	$< 10^{-4}$***	0.2698	0.0001***	0.0003***	0.065
	6	$< 10^{-4}$***	0.0341*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	7	$< 10^{-4}$***	0.5185	0.0002***	$< 10^{-4}$***	$< 10^{-4}$***
	8	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	9	$< 10^{-4}$***	0.0886	0.0027*	$< 10^{-4}$***	$< 10^{-4}$***
	10	$< 10^{-4}$***	0.0012*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***

*$p < 0.05$, uncorrected

**$p < 0.001$, uncorrected

***$p < 0.00037$, uncorrected (which is equivalent to $p < 0.05$, Bonferroni corrected)

and then conducting the global signal removal yielded a high accuracy of clustering in terms of the total GEV and WCSS. We obtained total GEV values that are somewhat larger than those obtained in previous studies for EEG microstate analysis [7, 9, 13, 66], which suggests that fMRI state-transition dynamics analysis may be promising. Furthermore, by carrying over the three clustering methods yielding the best clustering performance to a test–retest reliability analysis, we found that, for different fMRI data sets and different networks, test–retest reliability was higher in the within-participant comparison than the between-participant comparison. This result held true for most combinations of the number of clusters, $K \in \{2, \ldots, 10\}$, and index quantifying the estimated state-transition dynamics. We also found that the K-means clustering yielded the highest test–retest reliability among the three clustering methods. The present results suggest that clustering-based analysis of state-transition dynamics, which is substantially simpler than the hidden Markov model [12–14, 21–25, 27, 86] and the energy landscape analysis [28–31], may be a sufficiently competitive method to derive state-transition dynamics in fMRI data.

The microstate analysis was originally proposed for EEG data [5–8, 11, 87, 88]. Microstates in EEG data are typically of the order of 100 ms. One cannot directly associate the discrete states estimated from fMRI data with EEG or MEG microstates because the

Islam *et al. BMC Neuroscience*   (2024) 25:14

Page 15 of 19

Table 4 Results of the permutation test for the whole-brain network extracted from the HCP data

	K	Centroid	Coverage	Frequency	Lifespan	Trans. prob.
K-means	2	$< 10^{-4}$***	0.0308*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	3	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	4	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	5	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	6	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	7	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	8	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	9	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	10	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
TAAHC	2	$< 10^{-4}$***	0.0001***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	3	$< 10^{-4}$***	0.0016*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	4	$< 10^{-4}$***	0.0305*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	5	$< 10^{-4}$***	0.0297*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	6	$< 10^{-4}$***	0.0164*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	7	$< 10^{-4}$***	0.2864	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	8	$< 10^{-4}$***	0.0699	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	9	$< 10^{-4}$***	0.0203*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	10	$< 10^{-4}$***	0.0764	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
Bisecting K-means	2	$< 10^{-4}$***	0.0015*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	3	$< 10^{-4}$***	0.0531	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	4	$< 10^{-4}$***	0.0078*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	5	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	6	$< 10^{-4}$***	0.0009**	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	7	$< 10^{-4}$***	0.005*	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	8	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	9	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***
	10	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***	$< 10^{-4}$***

*$p < 0.05$, uncorrected

**$p < 0.001$, uncorrected

***$p < 0.00037$, uncorrected (which is equivalent to $p < 0.05$, Bonferroni corrected)

time resolution of fMRI data is much lower than 100 ms; a typical TR is approximately between 1 and 3 s. Furthermore, the typical duration of a discrete state is longer than one TR. For example, the average lifespan of a state was 3.3 TR, 2.5 TR, and 2.2 TR when we estimated four, seven, and ten states, respectively, for the DMN extracted from the MSC data. Therefore, cognitively or physiologically relevant discrete states estimated for fMRI data [19, 24, 67] may be different from those captured by microstates in EEG and MEG data. However, promising correspondences between EEG microstates and fMRI states have been reported [15, 19, 89, 90]. Analyzing simultaneously recorded EEG-fMRI data may further reveal connection between EEG microstates and discrete states for fMRI data [15, 18, 91–96].

We examined test–retest reliability of discrete states estimated by clustering activity pattern vectors of fMRI data. In contrast, various previous studies estimated discrete states by clustering functional networks from fMRI data [14, 19–21, 23, 25, 36–39]. Our methods of test–retest reliability analysis do not depend on how the discrete states are estimated and therefore are applicable to the case of state-transition dynamics of functional networks. To the best of our knowledge, no work has systematically compared the reliability between state-transition dynamics estimated for spatial activity patterns or their vectorized versions and those estimated

for functional networks from the same fMRI data. Such a comparative analysis may better inform us whether activity patterns or functional networks are more powerful biomarkers than the other when combined with state-transition dynamics modeling. In a similar vein, the aforementioned studies pursuing similarity between EEG microstates and fMRI dynamic states have been confined to the case in which fMRI dynamic states are estimated from dynamics of functional connectivity, not dynamics of activity patterns. These topics warrant future work.

In EEG microstate analysis, it is common to generate global microstate maps, which is to determine a given number of microstates by clustering candidate EEG maps obtained from different participants altogether. Then, one matches the obtained microstate maps shared by all the participants to the individual EEG maps from the individual participants to determine the microstate dynamics for each participant [9, 15, 66, 67, 97]. For EEG data, this approach has been shown to accrue higher reliability than microstate maps estimated separately for individual participants [9]. Nevertheless, we have estimated the states separately for individual participants (and for individual sessions) in the present study. This is because, for fMRI data, one often estimates state dynamics separately for each individual, which allows one to study subject variability of the estimated state dynamics or to exploit it [19, 21, 25, 98]. In contrast, pooling fMRI data from different participants to generate across-participant templates of discrete states is also a common practice [19, 32, 35, 67, 98]. In fact, one can run our test–retest reliability analysis even if we estimate the templates of the discrete states shared by all participants, with the exception of the cluster centroid, $\mathbf{c}_\ell$, as an observable of the estimated state dynamics; if we use a shared template, $\mathbf{c}_\ell$ is the same for all sessions and individuals and therefore one cannot compare its reliability within versus between participants. We point out that comparison of the reliability between shared templates and individualized templates of discrete states for fMRI data, as was done for EEG data [9], is underexplored.

We ran a permutation test to statistically compare the within-participant and between-participant test–retest reliability. This permutation test is an adaptation of what we recently developed for energy landscape analysis [50] to the case of clustering-based state-transition dynamics. This method is not limited to fMRI data. It is straightforward to use it for EEG and MEG microstate data analysis obtained from multiple participants and multiple sessions per participant. Our code is publicly available on GitHub [49]. The only requirement is to define observables and to be able to measure the discrepancy in the observable between an arbitrary pair of sessions. Assessing test–retest reliability in EEG [6,

9–11] and MEG [99, 100] data using this technique as well as furthering the application to fMRI data in health and disease may be fruitful.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12868-024-00854-3.

> **Additional file 1.** Supplementary information.

Availability of data and materials
Two publicly available data sets were used in this work. The first data set was provided by the Midnight Scan Club (MSC) project, funded by NIH Grants NS088590, TR000448 (NUFD), MH104592 (DJG), and HD087011 (to the Intellectual and Developmental Disabilities Research Center at Washington University); the Jacobs Foundation (NUFD); the Child Neurology Foundation (NUFD); the McDonnell Center for Systems Neuroscience (NUFD, BLS); the Mallinckrodt Institute of Radiology (NUFD); the Hope Center for Neurological Disorders (NUFD, BLS, SEP); and Dart Neuroscience LLC. This data was obtained from the OpenfMRI database. Its accession number is ds000224. The second data set was provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Code availability
The source code used for the computation of discrete state dynamics and the evaluation of their test–retest reliability used in this research paper can be accessed on the GitHub [49].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
All the authors have read and understood the BMC Neuroscience policy on declaration of interests and they declare that they have no competing interests.

References

1. Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. Proc Natl Acad Sci USA. 2001;98:676–82.
2. Rabinovich MI, Varona P, Selverston AI, Abarbanel HD. Dynamical principles in neuroscience. Rev Mod Phys. 2006;78:1213.
3. Deco G, Tononi G, Boly M, Kringelbach ML. Rethinking segregation and integration: contributions of whole-brain modelling. Nat Rev Neurosci. 2015;16:430–9.
4. Avena-Koenigsberger A, Misic B, Sporns O. Communication dynamics in complex brain networks. Nat Rev Neurosci. 2018;19:17–33.
5. Khanna A, Pascual-Leone A, Michel CM, Farzan F. Microstates in resting-state EEG: current status and future directions. Neurosci Biobehav Rev. 2015;49:105–13.
6. Michel CM, Koenig T. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review. NeuroImage. 2018;180:577–93.
7. Von Wegner F, Knaut P, Laufs H. EEG microstate sequences from different clustering algorithms are information-theoretically invariant. Front Comput Neurosci. 2018;12:70.
8. Koenig T, Prichep L, Lehmann D, Sosa PV, Braeker E, Kleinlogel H, Isenhart R, John ER. Millisecond by millisecond, year by year: normative EEG microstates and developmental stages. NeuroImage. 2002;16:41–8.
9. Khanna A, Pascual-Leone A, Farzan F. Reliability of resting-state microstate features in electroencephalography. PLoS ONE. 2014;9: 114163.
10. Liu J, Xu J, Zou G, He Y, Zou Q, Gao J-H. Reliability and individual specificity of EEG microstate characteristics. Brain Topogr. 2020;33:438–49.
11. Zhang K, Shi W, Wang C, Li Y, Liu Z, Liu T, Li J, Yan X, Wang Q, Cao Z, et al. Reliability of EEG microstate analysis at different electrode densities during propofol-induced transitions of brain states. NeuroImage. 2021;231: 117861.
12. Tait L, Zhang J. +microstate: a MATLAB toolbox for brain microstate analysis in sensor and cortical EEG/MEG. NeuroImage. 2022;258: 119346.
13. Tait L, Zhang J. MEG cortical microstates: spatiotemporal characteristics, dynamic functional connectivity and stimulus-evoked responses. NeuroImage. 2022;251: 119006.
14. Baker AP, Brookes MJ, Rezek IA, Smith SM, Behrens T, Probert Smith PJ, Woolrich M. Fast transient networks in spontaneous human brain activity. eLife. 2014;3: e01867.
15. Britz J, Van De Ville D, Michel CM. Bold correlates of EEG topography reveal rapid resting-state network dynamics. NeuroImage. 2010;52:1162–70.
16. Schwab S, Koenig T, Morishima Y, Dierks T, Federspiel A, Jann K. Discovering frequency sensitive thalamic nuclei from EEG microstate informed resting state fMRI. NeuroImage. 2015;118:368–75.
17. Case M, Zhang H, Mundahl J, Datta Y, Nelson S, Gupta K, He B. Characterization of functional brain activity and connectivity using EEG and fMRI in patients with sickle cell disease. NeuroImage Clin. 2017;14:1–17.
18. Bréchet L, Brunet D, Birot G, Gruetter R, Michel CM, Jorge J. Capturing the spatiotemporal dynamics of self-generated, task-initiated thoughts with EEG and fMRI. NeuroImage. 2019;194:82–92.
19. Allen EA, Damaraju E, Plis SM, Erhardt EB, Eichele T, Calhoun VD. Tracking whole-brain connectivity dynamics in the resting state. Cereb Cortex. 2014;24:663–76.
20. Calhoun VD, Miller R, Pearlson G, Adalı T. The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. Neuron. 2014;84:262–74.
21. Nielsen SF, Schmidt MN, Madsen KH, Mørup M. Predictive assessment of models for dynamic functional connectivity. NeuroImage. 2018;171:116–34.
22. Ezaki T, Himeno Y, Watanabe T, Masuda N. Modelling state-transition dynamics in resting-state brain signals by the hidden Markov and Gaussian mixture models. Eur J Neurosci. 2021;54:5404–16.
23. Ryali S, Supekar K, Chen T, Kochalka J, Cai W, Nicholas J, Padmanabhan A, Menon V. Temporal dynamics and developmental maturation of salience, default and central-executive network interactions revealed by variational Bayes hidden Markov modeling. PLoS Comput Biol. 2016;12:1005138.
24. Vidaurre D, Smith SM, Woolrich MW. Brain network dynamics are hierarchically organized in time. Proc Natl Acad Sci USA. 2017;114:12827–32.
25. Taghia J, Ryali S, Chen T, Supekar K, Cai W, Menon V. Bayesian switching factor analysis for estimating time-varying functional connectivity in fMRI. NeuroImage. 2017;155:271–90.
26. Brookes MJ, Groom MJ, Liuzzi L, Hill RM, Smith HJ, Briley PM, Hall EL, Hunt BA, Gascoyne LE, Taylor MJ, et al. Altered temporal stability in dynamic neural networks underlies connectivity changes in neurodevelopment. NeuroImage. 2018;174:563–75.
27. Vidaurre D. A new model for simultaneous dimensionality reduction and time-varying functional connectivity estimation. PLoS Comput Biol. 2021;17:1008580.
28. Watanabe T, Hirose S, Wada H, Imai Y, Machida T, Shirouzu I, Konishi S, Miyashita Y, Masuda N. Energy landscapes of resting-state brain networks. Front Neuroinform. 2014;8:12.
29. Watanabe T, Masuda N, Megumi F, Kanai R, Rees G. Energy landscape and dynamics of brain activity during human bistable perception. Nat Commun. 2014;5:4765.
30. Ezaki T, Watanabe T, Ohzeki M, Masuda N. Energy landscape analysis of neuroimaging data. Philos Trans R Soc A. 2017;375:20160287.
31. Ezaki T, Sakaki M, Watanabe T, Masuda N. Age-related changes in the ease of dynamical transitions in human brain activity. Hum Brain Mapp. 2018;39:2673–88.
32. Liu X, Chang C, Duyn JH. Decomposition of spontaneous brain activity into distinct fMRI co-activation patterns. Front Syst Neurosci. 2013;7:101.
33. Liu X, Duyn JH. Time-varying functional network information extracted from brief instances of spontaneous brain activity. Proc Natl Acad Sci USA. 2013;110:4392–7.
34. Liu X, Zhang N, Chang C, Duyn JH. Co-activation patterns in resting-state fMRI signals. NeuroImage. 2018;180:485–94.
35. Paakki J-J, Rahko JS, Kotila A, Mattila M-L, Miettunen H, Hurtig TM, Jussila KK, Kuusikko-Gauffin S, Moilanen IK, Tervonen O, et al. Co-activation pattern alterations in autism spectrum disorder—a volume-wise hierarchical clustering fMRI study. Brain Behav. 2021;11:02174.
36. Sakoğlu Ü, Pearlson GD, Kiehl KA, Wang YM, Michael AM, Calhoun VD. A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. Magn Reson Mater Phys Biol Med. 2010;23:351–66.
37. Hutchison RM, Womelsdorf T, Gati JS, Everling S, Menon RS. Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. Hum Brain Mapp. 2013;34:2154–77.
38. Leonardi N, Richiardi J, Gschwind M, Simioni S, Annoni J-M, Schluep M, Vuilleumier P, Van De Ville D. Principal components of functional connectivity: a new approach to study dynamic brain connectivity during rest. NeuroImage. 2013;83:937–50.
39. Abrol A, Chaze C, Damaraju E, Calhoun VD. The chronnectome: evaluating replicability of dynamic connectivity patterns in 7500 resting fMRI datasets. In: 2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE; 2016. p. 5571–74.
40. Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, Papademetris X, Constable RT. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat Neurosci. 2015;18:1664–71.
41. Gordon EM, Laumann TO, Gilmore AW, Newbold DJ, Greene DJ, Berg JJ, Ortega M, Hoyt-Drazen C, Gratton C, Sun H, et al. Precision functional mapping of individual human brains. Neuron. 2017;95:791–807.
42. Amico E, Goñi J. The quest for identifiability in human functional connectomes. Sci Rep. 2018;8:8254.
43. Noble S, Scheinost D, Constable RT. A decade of test–retest reliability of functional connectivity: a systematic review and meta-analysis. NeuroImage. 2019;203: 116157.
44. Venkatesh M, Jaja J, Pessoa L. Comparing functional connectivity matrices: a geometry-aware approach applied to participant identification. NeuroImage. 2020;207: 116398.
45. Chiêm B, Abbas K, Amico E, Duong-Tran DA, Crevecoeur F, Goñi J. Improving functional connectome fingerprinting with degree-normalization. Brain Connect. 2022;12:180–92.

Islam *et al. BMC Neuroscience*   (2024) 25:14

Page 18 of 19

46. Zhang C, Baum SA, Adduru VR, Biswal BB, Michael AM. Test–retest reliability of dynamic functional connectivity in resting state fMRI. NeuroImage. 2018;183:907–18.

47. Zhang X, Liu J, Yang Y, Zhao S, Guo L, Han J, Hu X. Test–retest reliability of dynamic functional connectivity in naturalistic paradigm functional magnetic resonance imaging. Hum Brain Mapp. 2022;43:1463–76.

48. Long Y, Ouyang X, Yan C, Wu Z, Huang X, Pu W, Cao H, Liu Z, Palaniyappan L. Evaluating test–retest reliability and sex-/age-related effects on temporal clustering coefficient of dynamic functional brain networks. Hum Brain Mapp. 2023;44:2191–208.

49. Islam S. Functional MRI state transition dynamics, GitHub repository. 2023. https://github.com/sislam99/fmri_state_transition_dynamics. Accessed 5 Aug 2023.

50. Khanra P, Nakuci J, Muldoon S, Watanabe T, Masuda N. Reliability of energy landscape analysis of resting-state functional MRI data. arXiv preprint. 2023. arXiv:2305.19573.

51. Fair DA, Cohen AL, Power JD, Dosenbach NU, Church JA, Miezin FM, Schlaggar BL, Petersen SE. Functional brain networks develop from a "local to distributed" organization. PLoS Comput Biol. 2009;5:1000381.

52. Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, et al. Functional network organization of the human brain. Neuron. 2011;72:665–78.

53. Watanabe T, Rees G. Brain network dynamics in high-functioning individuals with autism. Nat Commun. 2017;8:1–14.

54. Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, WU-Minn HCP Consortium. The WU-Minn Human Connectome Project: an overview. NeuroImage. 2013;80:62–79.

55. Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR, et al. The minimal preprocessing pipelines for the human connectome project. NeuroImage. 2013;80:105–24.

56. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage. 2002;17:825–41.

57. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage. 2012;59:2142–54.

58. Burgess GC, Kandala S, Nolan D, Laumann TO, Power JD, Adeyemo B, Harms MP, Petersen SE, Barch DM. Evaluation of denoising strategies to address motion-correlated artifacts in resting-state functional magnetic resonance imaging data from the human connectome project. Brain Connect. 2016;6:669–80.

59. Siegel JS, Mitra A, Laumann TO, Seitzman BA, Raichle M, Corbetta M, Snyder AZ. Data quality influences observed links between functional connectivity and behavior. Cereb Cortex. 2017;27:4492–502.

60. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage. 2002;15:273–89.

61. Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo X-N, Holmes AJ, Eickhoff SB, Yeo BT. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cereb Cortex. 2018;28:3095–114.

62. Murphy K, Fox MD. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. NeuroImage. 2017;154:169–73.

63. Li J, Bolt T, Bzdok D, Nomi JS, Yeo B, Spreng RN, Uddin LQ. Topography and behavioral relevance of the global signal in the human brain. Sci Rep. 2019;9:14286.

64. Aquino KM, Fulcher BD, Parkes L, Sabaroedin K, Fornito A. Identifying and removing widespread signal deflections from fMRI data: rethinking the global signal regression problem. NeuroImage. 2020;212: 116614.

65. Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE. Methods to detect, characterize, and remove motion artifact in resting state fMRI. NeuroImage. 2014;84:320–41.

66. Zanesco AP, King BG, Skwara AC, Saron CD. Within and between-person correlates of the temporal dynamics of resting EEG microstates. NeuroImage. 2020;211: 116631.

67. Abrol A, Damaraju E, Miller RL, Stephen JM, Claus ED, Mayer AR, Calhoun VD. Replicability of time-varying connectivity patterns in large resting state fMRI samples. NeuroImage. 2017;163:160–76.

68. Khosla M, Jamison K, Ngo GH, Kuceyeski A, Sabuncu MR. Machine learning in resting-state fMRI analysis. Magn Reson Imaging. 2019;64:101–21.

69. Arthur D, Vassilvitskii S. K-means++ the advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms. 2007. p. 1027–35.

70. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825–30.

71. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. Hoboken: Wiley; 2009.

72. Tibshirani R, Walther G. Cluster validation by prediction strength. J Comput Graph Stat. 2005;14:511–28.

73. Murray MM, Brunet D, Michel CM. Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr. 2008;20:249–64.

74. Poulsen AT, Pedroni A, Langer N, Hansen LK. Microstate EEGlab toolbox: an introductory guide. BioRxiv. 2018. https://doi.org/10.1101/289850.

75. Pascual-Marqui RD, Michel CM, Lehmann D. Segmentation of brain electrical activity into microstates: model estimation and validation. IEEE Trans Biomed Eng. 1995;42:658–65.

76. Steinbach M, Karypis G, Kumar V. A comparison of document clustering techniques. Technical report 00-034, Department of Computer Science and Egineering, University of Minnesota (2000). https://hdl.handle.net/11299/215421. Accessed 04 Oct 2022.

77. Bishop CM, Nasrabadi NM. Pattern recognition and machine learning, vol. 4. New York: Springer; 2006.

78. Lindsay BG. Mixture models: theory, geometry, and applications. In: NSF-CBMS regional conference series in probability and statistics, vol. 5. Institute of Mathematical Statistics, The United States of America; 1995. p. 163.

79. Jain AK, Dubes RC. Algorithms for clustering data. Englewood Cliffs, NJ: Prentice-Hall Inc; 1988.

80. Bernardi S, Benna MK, Rigotti M, Munuera J, Fusi S, Salzman CD. The geometry of abstraction in the hippocampus and prefrontal cortex. Cell. 2020;183:954–67.

81. Liu Y, Brincat SL, Miller EK, Hasselmo ME. A geometric characterization of population coding in the prefrontal cortex and hippocampus during a paired-associate learning task. J Cogn Neurosci. 2020;32:1455–65.

82. Maris E, Oostenveld R. Nonparametric statistical testing of EEG-and MEG-data. J Neurosci Methods. 2007;164:177–90.

83. Zhang J, Northoff G. Beyond noise to function: reframing the global brain activity and its dynamic topography. Commun Biol. 2022;5:1350.

84. Wang Y, Yang C, Li G, Ao Y, Jiang M, Cui Q, Pang Y, Jing X. Frequency-dependent effective connections between local signals and the global brain signal during resting-state. Cogn Neurodyn. 2023;17:555–60.

85. Xia M, Wang J, He Y. BrainNet viewer: a network visualization tool for human brain connectomics. PLoS ONE. 2013;8:68910.

86. Vidaurre D, Hunt LT, Quinn AJ, Hunt BA, Brookes MJ, Nobre AC, Woolrich MW. Spontaneous cortical activity transiently organises into frequency specific phase-coupling networks. Nat Commun. 2018;9:2987.

87. Lehmann D, Ozaki H, Pal I. EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. Electroencephalogr Clin Neurophysiol. 1987;67:271–88.

88. Hatz F, Hardmeier M, Bousleiman H, Rüegg S, Schindler C, Fuhr P. Reliability of functional connectivity of electroencephalography applying microstate-segmented versus classical calculation of phase lag index. Brain Connect. 2016;6:461–9.

89. Yuan H, Zotev V, Phillips R, Drevets WC, Bodurka J. Spatiotemporal dynamics of the brain at rest-exploring EEG microstates as electrophysiological signatures of bold resting state networks. NeuroImage. 2012;60:2062–72.

90. Abreu R, Jorge J, Leal A, Koenig T, Figueiredo P. EEG microstates predict concurrent fMRI dynamic functional connectivity states. Brain Topogr. 2021;34:41–55.

91. Chang C, Liu Z, Chen MC, Liu X, Duyn JH. EEG correlates of time-varying bold functional connectivity. NeuroImage. 2013;72:227–36.

92. Preti MG, Leonardi N, Karahanoğlu FI, Grouiller F, Genetti M, Seeck M, Vulliemoz S, Van De Ville D. Epileptic network activity revealed by dynamic functional connectivity in simultaneous EEG-fMRI. In: 2014 IEEE 11th international symposium on biomedical imaging (ISBI). IEEE; 2014. p. 9–12.

93. Ahmad RF, Malik AS, Kamel N, Reza F, Abdullah JM. Simultaneous EEG-fMRI for working memory of the human brain. Australas Phys Eng Sci Med. 2016;39:363–78.

94. Keinänen T, Rytky S, Korhonen V, Huotari N, Nikkinen J, Tervonen O, Palva JM, Kiviniemi V. Fluctuations of the EEG-fMRI correlation reflect intrinsic strength of functional connectivity in default mode network. J Neurosci Res. 2018;96:1689–98.

95. Lurie DJ, Kessler D, Bassett DS, Betzel RF, Breakspear M, Kheilholz S, Kucyi A, Liégeois R, Lindquist MA, McIntosh AR, et al. Questions and controversies in the study of time-varying functional connectivity in resting fMRI. Netw Neurosci. 2020;4:30–69.

96. Mulert C. Simultaneous EEG and fMRI: towards the characterization of structure and dynamics of brain networks. Dialogues Clin Neurosci. 2013;15:381–6.

97. Lehmann D, Faber PL, Galderisi S, Herrmann WM, Kinoshita T, Koukkou M, Mucci A, Pascual-Marqui RD, Saito N, Wackermann J, et al. EEG micro-state duration and syntax in acute, medication-Naive, first-episode schizophrenia: a multi-center study. Psychiatry Res NeuroImaging. 2005;138:141–56.

98. Rashid B, Damaraju E, Pearlson GD, Calhoun VD. Dynamic connectivity states estimated from resting fMRI identify differences among Schizophrenia, bipolar disorder, and healthy control subjects. Front Hum Neurosci. 2014;8:897.

99. Garcés P, Martín-Buro MC, Maestú F. Quantifying the test–retest reliability of magnetoencephalography resting-state functional connectivity. Brain Connect. 2016;6:448–60.

100. Candelaria-Cook FT, Schendel ME, Ojeda CJ, Bustillo JR, Stephen JM. Reduced parietal alpha power and psychotic symptoms: test–retest reliability of resting-state magnetoencephalography in schizophrenia and healthy controls. Schizophr Res. 2020;215:229–40.

Publisher's Note