



OPEN

# Revisiting the global workspace orchestrating the hierarchical organization of the human brain

**Gustavo Deco** <sup>1,2,3,4</sup> , **Diego Vidaurre** <sup>5,6</sup> and **Morten L. Kringelbach** <sup>5,7,8</sup>

**A central challenge in neuroscience is how the brain organizes the information necessary to orchestrate behaviour. Arguably, this whole-brain orchestration is carried out by a core subset of integrative brain regions, a 'global workspace', but its constitutive regions remain unclear. We quantified the global workspace as the common regions across seven tasks as well as rest, in a common 'functional rich club'. To identify this functional rich club, we determined the information flow between brain regions by means of a normalized directed transfer entropy framework applied to multimodal neuroimaging data from 1,003 healthy participants and validated in participants with retest data. This revealed a set of regions orchestrating information from perceptual, long-term memory, evaluative and attentional systems. We confirmed the causal significance and robustness of our results by systematically lesioning a generative whole-brain model. Overall, this framework describes a complex choreography of the functional hierarchical organization of the human brain.**

Over recent decades, research within systems neuroscience has suggested that the brain is hierarchically organized in terms of anatomical structure<sup>1–7</sup> and function<sup>8–12</sup>. Nevertheless, a key remaining challenge is to determine precisely how this hierarchical organization allows the brain to orchestrate function and behaviour by organizing the flow of information and the underlying computations necessary for survival.

A large body of research has argued that whole-brain orchestration is likely to be carried out by a core subset of integrative brain regions. For example, according to the classic model of Norman and Shallice<sup>13</sup>, this processing involves the prefrontal cortices in charge of the supervisory attentional regulation of lower-level sensorimotor chains. In contrast, Baars proposed the concept of a 'global workspace' (GW), where information is integrated in a small group of brain regions before being broadcast to many other regions across the whole brain<sup>14</sup>. Extending this framework, Dehaene and Changeux<sup>15</sup> proposed their 'global neuronal workspace' hypothesis that associative perceptual, motor, attention, memory and value areas interconnect to form a higher-level unified space where information is broadly shared and broadcast back to lower-level processors. Colloquially, the GW is thus akin to a small core assembly of people in charge of an organization. Larger brain network organization has been shown to be efficient, robust and largely fault tolerant<sup>4,16,17</sup>, yet the effects of lesioning a core assembly—like the GW—are currently unknown.

Until recently, a key obstacle to advancing our understanding of the human brain's functional hierarchical organization has been the lack of suitable whole-brain measurements. However, the advent of Big Data, such as the Human Connectome Project (HCP)<sup>18,19</sup>, has created large multimodal whole-brain neuroimaging datasets of healthy individuals both in resting state and whilst performing many different tasks. The development of more advanced neuroimaging methods could allow for the estimation of

the bi-directional flow of information between all regions across the whole brain, which could subsequently be used to characterize the functional hierarchical organization of the brain. Many methods exist to estimate the information flow using different neuroimaging modalities, which have greatly contributed to our current knowledge (for example, refs. <sup>20–22</sup>). Here, we use this method to address the key challenge of identifying the functional hierarchical organization of the brain.

The main aim of our study is to characterize the GW as the core set of brain regions responsible for integration and orchestration. To do this, we have to determine the functional hierarchical whole-brain organization. For this purpose, we propose the concept of a 'functional rich club' (FRIC) as the core set of regions, an array of functional hubs that are characterized by a tendency to be more densely functionally connected among themselves than to other brain regions from where they receive integrative information. This notion is related to previous static descriptions of the anatomical rich club<sup>5,7</sup>, which includes nodes in a network with a tendency for high-degree nodes to be more densely connected among themselves than nodes of a lower degree. However, unlike the anatomical rich club, FRIC is a dynamic measure based on bidirectional flow of information not constrained by anatomy and thus will change across different tasks. Using our colloquial example of a core assembly, for different tasks some people remain through all executive meetings while others are substituted in and out on the basis of their expertise. In a similar manner, FRIC would include both common and task-specific brain regions as a result of the different flows of information for different kinds of task. Following the original ideas of Baars<sup>14</sup>, we propose that the invariant 'GW' is the intersection of the different sets of task-related FRICs.

To achieve our goal of identifying the GW as the intersection of FRICs from data of different tasks and in resting condition, we develop a normalized directed transfer entropy (NDTE) framework.

<sup>1</sup>Center for Brain and Cognition, Computational Neuroscience Group, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain. <sup>2</sup>Institut de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. <sup>3</sup>Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. <sup>4</sup>School of Psychological Sciences, Monash University, Clayton, Victoria, Australia. <sup>5</sup>Department of Psychiatry, University of Oxford, Oxford, UK. <sup>6</sup>Center for Functionally Integrative Neuroscience, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark. <sup>7</sup>Centre for Eudaimonia and Human Flourishing, University of Oxford, Oxford, UK. <sup>8</sup>Center for Music in the Brain, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark. e-mail: [gustavo.deco@upf.edu](mailto:gustavo.deco@upf.edu); [morten.kringelbach@psych.ox.ac.uk](mailto:morten.kringelbach@psych.ox.ac.uk)

This framework provides a bidirectional description of the functional information flow underlying brain signals from 1,003 HCP participants in resting state and when carrying out seven tasks.

Finally, we validate our finding of the invariant GW through constructing and selectively lesioning a whole-brain model that accurately simulates the empirical functional hierarchy and thus describes the underlying dynamical mechanisms. Systematic lesioning of subsets of regions in the FRIC in this model establishes the generative role of the GW in orchestrating function and allows us to characterize its efficiency and robustness.

## Results

Figure 1 sketches the overall aim of finding the regions orchestrating the functional hierarchical organization of the brain, sometimes called the ‘global workspace’<sup>14,15</sup>.

**Functional hierarchical organization of resting.** We characterized the functional hierarchical organization of the resting state of 1,003 participants. For this purpose, we extracted the bidirectional flow of information between brain regions using the concept of NDTE (see Methods), which is an information-theoretical measure of causality between two time series. This allows us to infer the underlying bidirectional reciprocal communication between any source and target regions. Specifically, first we compute the mutual information directed flow, that is the predictability of a target in the future given the past of the source region, beyond the predictability from its own past (see Eq. 1 in Methods). Then the mutual information directed flow is normalized by the mutual information that both source and target have about the future of the target (see Eq. 8 in Methods). At the individual level, we compute the statistical significance by using the circular time-shifted surrogates method, which has been shown to be particularly well adapted to causal measurements<sup>23</sup>. At the group level, we aggregate the *P* values corresponding to each pairwise NDTE flow using the Stouffer method<sup>24</sup> and correct for multiple comparisons (see Methods).

Importantly, the NDTE framework was thoroughly validated through network simulations (see Methods and Supplementary materials). Furthermore, in the Methods and Discussion, we show how other research has validated the transfer entropy framework in neuroimaging<sup>25</sup> and the relevance of using circular shift surrogates for eliminating spurious inferences<sup>26,27</sup>.

We computed a matrix containing the NDTE flow between the regions in four different parcellations (see Methods). For a fine-scale parcellation, we used a modified version of the Glasser parcellation with a total of 378 regions (360 cortical and 18 subcortical regions)<sup>28</sup>. For a medium-scale parcellation, we used the Brainnetome parcellation with a total of 246 regions (210 cortical and 36 subcortical regions)<sup>29</sup>. For a coarser-scale parcellation suitable for whole-brain computational modelling, we used a modified Desikan–Killiany parcellation which included subcortical regions (DK80, 62 cortical regions and 18 subcortical regions)<sup>30,31</sup>.

To establish the hierarchical organization, we compute the total incoming and outgoing information for all brain regions. More specifically, for each brain region, the total incoming flow of information,  $G_{in}$ , is the sum of all sources (that is, the sum over the rows in the matrix). Similarly, the total outgoing flow of information,  $G_{out}$ , is the sum over all targets (that is, the sum over the columns). We also compute the total information being processed in a brain region as the sum:  $G_{tot} = G_{in} + G_{out}$ , which is possible given that the measures have been normalized (see above and Methods).

We show the functional hierarchy described by each of the incoming ( $G_{in}$ ), outgoing ( $G_{out}$ ) and total ( $G_{tot}$ ) directional information flow computed from the NDTE matrix of 1,003 participants (see Fig. 1 and Methods) in the Glasser (Fig. 2a) and DK80 parcellations (Fig. 2d). As can be seen (Extended Data Fig. 1), the outgoing information flow,  $G_{out}$ , is highest in sensory areas, while the

incoming information,  $G_{in}$ , is highest in higher-order, integrative transmodal areas.

On the other hand, a popular proxy for anatomical hierarchy is the myelination of brain regions as measured by myelin-weighted T1w/T2w<sup>32</sup>. It should be noted that there is controversy in the literature as to whether the ratio of T1w and T2w images (T1w/T2w ratio) in grey matter is a reliable estimate of cortical myelin content. Nevertheless, there is a growing consensus that the T1w/T2w ratio is a marker of general white-matter microstructure in both health and disease<sup>33–35</sup> provided that a correctly standardized measure of T1w/T2w is applied<sup>36</sup>. Our analysis uses the neuroimaging in the HCP dataset of healthy participants with isotropic spatial resolution of 0.7 mm, which enables highly accurate individual maps of cortical myelin content and thickness<sup>18</sup>.

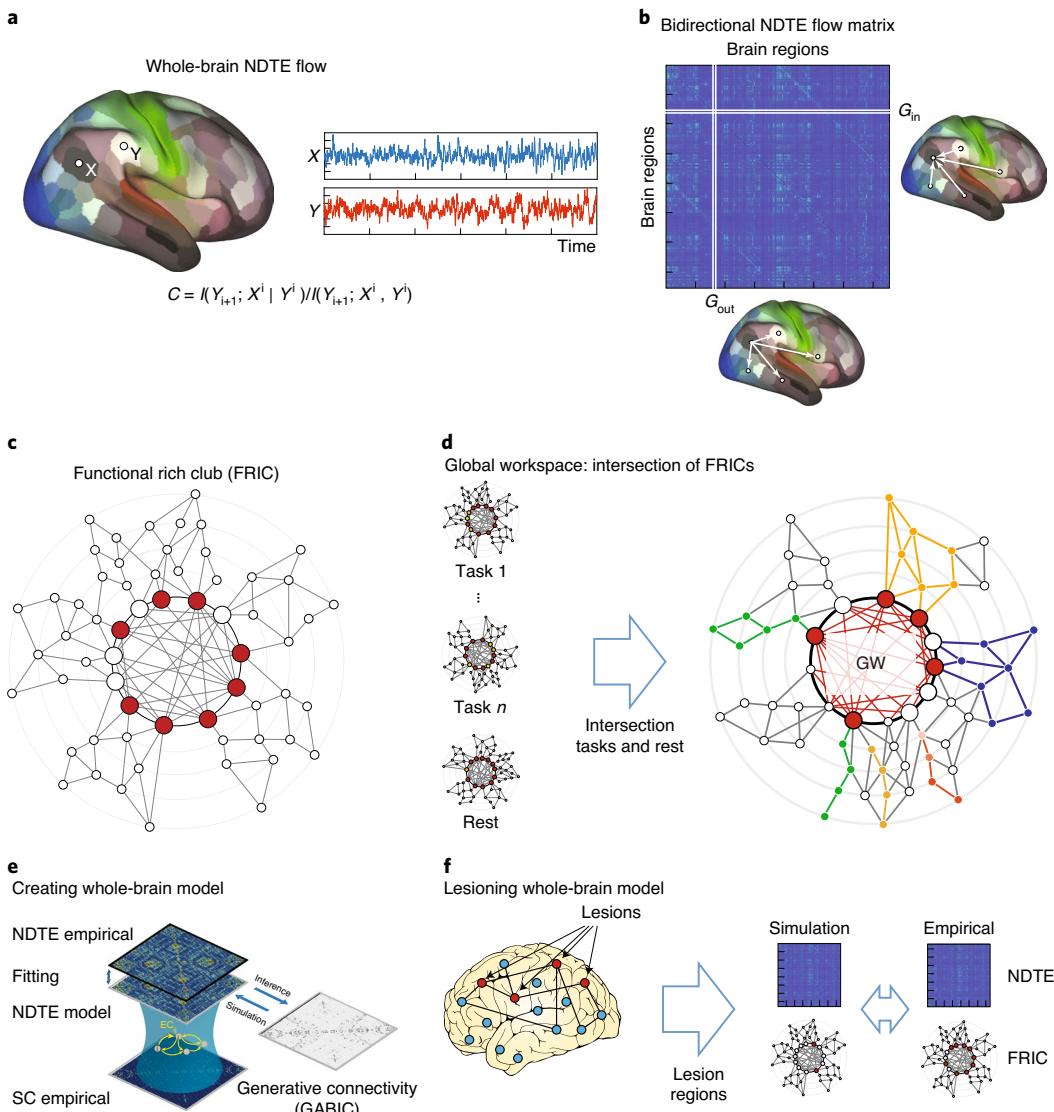
We render the T1w/T2w in the Glasser (Fig. 2b and Supplementary Fig. 1) and DK80 parcellations (Fig. 2e and Supplementary Fig. 1). Important information about the driving nature of sensory areas (more myelin) and the integrative role of higher-order transmodal areas (less myelin) has been demonstrated from this structural information in recent papers<sup>8,37,38</sup>. However, this structural measure does not change with different tasks and is therefore unlikely to capture the functional dynamic changes in hierarchical organization.

We provide scatterplots between the functional hierarchy ( $G_{in}$ ,  $G_{out}$  and  $G_{tot}$ ) and structural hierarchy (myelination) for the cortical regions in the Glasser (Fig. 2c) and DK80 parcellations (Fig. 2f). The linear Pearson correlations are shown by the red line (with standard error in shaded gray) overlaid on the scatterplots, with values for the Glasser parcellation of  $r(720) = 0.29$  for  $G_{in}$  ( $P < 0.001$ , 95% CI 0.19–0.38),  $r(720) = 0.57$  for  $G_{out}$  ( $P < 0.001$ , 95% CI = 0.50–0.64) and  $r(720) = 0.63$  for  $G_{tot}$  ( $P < 0.001$ , 95% CI = 0.56–0.69) and values for the DK80 parcellation of  $r(124) = 0.07$  for  $G_{in}$  (not significant, 95% CI –0.18 to 0.32),  $r(124) = 0.53$  for  $G_{out}$  ( $P < 0.001$ , 95% CI 0.32–0.69) and  $r(124) = 0.61$  for  $G_{tot}$  ( $P < 0.001$ , 95% CI 0.42–0.74). This shows that myelination is highly correlated with  $G_{tot}$ , and mainly driven by correlation with the outgoing flow,  $G_{out}$ . Importantly, as expected, the level of correlation between function and structure is lower for the incoming flow (the integrative measure of  $G_{in}$ ). This is consistent with the demonstration of a direct link between electrophysiological connectivity and myelination<sup>39</sup>.

To validate the results from functional magnetic resonance imaging (fMRI), we also characterized the functional hierarchical organization for the corresponding HCP magnetoencephalography (MEG) time series in the 62 cortical regions of the DK80 parcellation. Attesting to the robustness of the results, we found similar correlations between  $G_{in}$  and  $G_{out}$  from HCP MEG data and fMRI:  $r(124) = 0.40$  for  $G_{in}$  (14–22 Hz, window size 1,000 ms,  $P < 0.001$ , 95% CI 0.17–0.59),  $r(124) = 0.41$  for  $G_{out}$  (22.5–30.5 Hz, window size 500 ms,  $P < 0.001$ , 95% CI 0.18–0.60) (see Extended Data Fig. 2b). HCP MEG data and T1w/T2w:  $r(124) = 0.04$  for  $G_{in}$  (14–22 Hz, window size 1,000 ms, not significant, 95% CI –0.21 to 0.29),  $r(124) = 0.48$  for  $G_{out}$  (22.5–30.5 Hz, window size 500 ms,  $P < 0.001$ , 95% CI 0.26–0.65) (see Extended Data Fig. 2a).

**Functional hierarchy across different tasks and rest.** These results show that the measures of  $G_{in}$  and  $G_{out}$  are very different, and we quantified the changes in the relationship between the functional hierarchy across different tasks and rest. We used all seven tasks in the HCP dataset. The task battery was designed to cover as wide a range of neural systems as is feasible within realistic time constraints<sup>40</sup> (see Methods).

Figure 3a provides cortical renderings of all seven tasks and rest for the incoming ( $G_{in}$ ), outgoing ( $G_{out}$ ) and total ( $G_{tot}$ ) directional information flow computed from the NDTE matrix of participants (see Fig. 1 and Methods) in the DK80 parcellation (using 3D views from the side and midline). Importantly, we validated the NDTE framework on 45 HCP participants with retest data (Extended Data



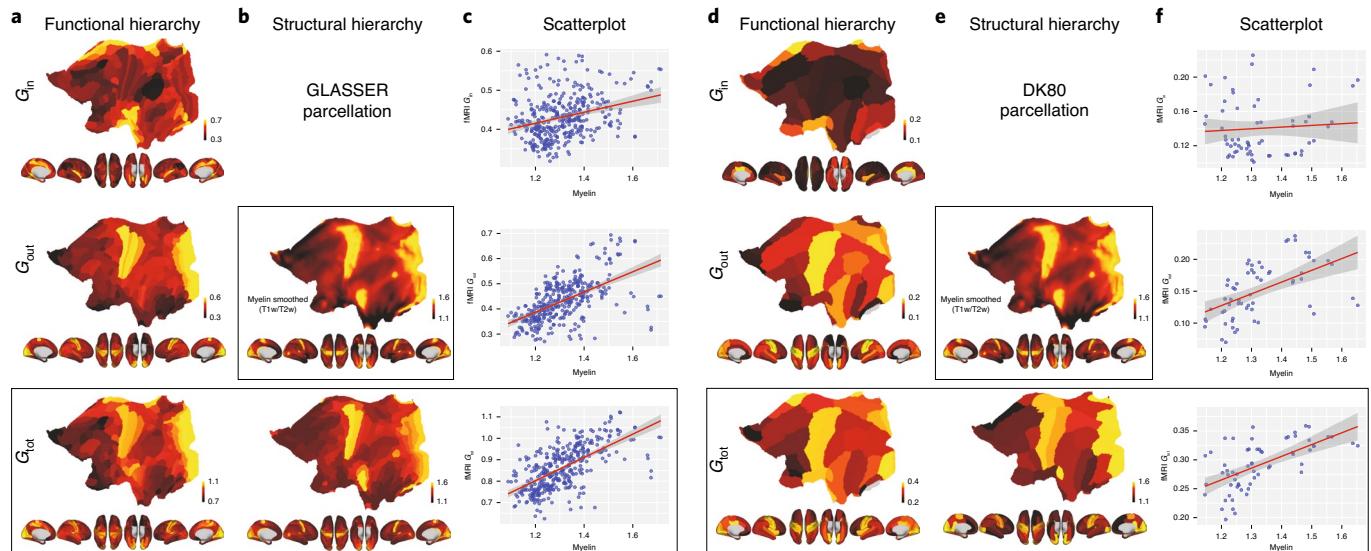
**Fig. 1 | Overview of general theoretical framework.** **a**, The causal bidirectional flow of information between any two brain regions is determined by computing the pairwise NDTE. The figure shows two arbitrary regions, X and Y, superimposed on the human brain, with different shades of colours marking different regions, and we show the timecourses of X and Y in blue and red, respectively. The statistical significance is determined at the individual level using the circular time-shifted surrogates method<sup>23</sup> and at the group level using  $P$  level aggregation across individuals. **b**, The functional hierarchical organization is given by the full NDTE matrix, where the rows contain the target regions and the columns contain the source regions. For each brain region, the total incoming flow of information,  $G_{in}$ , is simply the sum of all sources (that is, the sum over the rows in the matrix). Similarly, the total outgoing flow of information,  $G_{out}$ , is the sum over all targets (that is, the sum over the columns). **c**, The FRIC is the smallest set of brain regions (highlighted in red) that integrate and orchestrate function in a given task. It can be identified as the most highly connected brain regions that (1) are more densely connected within themselves than to regions with lower connectivity, whilst (2) having the highest level of incoming directed flow ( $G_{in}$ ) and (3) the lowest outgoing directed flow ( $G_{out}$ ; see Methods). **d**, The global neuronal workspace must be relevant to all tasks and situations (highlighted with red circles for each task) and must therefore comprise the common FRIC members across many different tasks, that is, the intersection of FRICs from tasks and rest (highlighted with red circles on the right panel, which also shows networks lower in the hierarchy in different colours). **e**, To establish the causal importance of the FRIC, we fit a whole-brain model to the resting NDTE empirical data and extract the underlying effective connectivity (see Results and Methods). **f**, The whole-brain model is then systematically lesioned for regions belonging to the FRIC and compared with lesioning non-FRIC members. For illustration, the cartoon highlights the lesion sites with red dots and other regions with blue dots. Overall, this confirms the causal importance of these regions in the orchestration of the functional hierarchical organization of the human brain.

Fig. 3) and found very high reproducibility (correlation up to a 0.97) despite using only a subsample of 45 of the 1,003 participants.

We also computed and rendered the NDTE matrices in the finer Brainnetome246 parcellation for all seven tasks and rest for the incoming ( $G_{in}$ ), outgoing ( $G_{out}$ ) and total ( $G_{tot}$ ) directional information flow for all participants (see Extended Data Fig. 4 and

Methods). The results are very similar and consistent across tasks and rest in the two parcellations.

Quantification of the differences is provided by the correlation matrices between incoming ( $G_{in}$ ), outgoing ( $G_{out}$ ) and total ( $G_{tot}$ ) directional information flow in the seven tasks and rest. The results show that there are significant differences between the hierarchy



**Fig. 2 | Comparing functional and structural hierarchical organization.** **a**, Functional hierarchy shown as cortical renderings of each of the incoming ( $G_{in}$ ), outgoing ( $G_{out}$ ) and total ( $G_{tot}$ ) directional information flow computed from the NDTE matrix of 1,003 HCP participants (see Fig. 1 and Methods) in the Glasser parcellation (using renderings of cortical flattening and 3D renderings with midline, right, left, top and bottom views). As shown, the outgoing information,  $G_{out}$ , is highest in sensory areas, while the incoming information,  $G_{in}$ , is highest in higher-order, integrative transmodal areas. **b**, The structural hierarchy shown for myelination of brain regions (myelin-weighted T1w/T2w) with the same renderings at the voxel level (top) and in the Glasser parcellation (bottom). **c**, Scatterplots between the functional hierarchy ( $G_{in}$ ,  $G_{out}$  and  $G_{tot}$ ) and structural hierarchy (myelination). The linear correlations are shown by the red line (with standard error in shaded gray) overlaid on the scatterplots (see Results for  $r$  values). This shows that myelination is highly correlated with  $G_{tot}$  and mainly driven by correlation with  $G_{out}$ . On the other hand, there is a much lower correlation with the incoming flow, that is, an integrative measure of  $G_{in}$ . This means that the static measure of myelination is likely to mostly reflect the driving flow in sensory areas, but provides much less information on integrative areas. **d**, The same as **a** but for the DK80 parcellation. **e**, Myelination in the DK80 parcellation. **f**, Scatterplots between the functional hierarchy ( $G_{in}$ ,  $G_{out}$  and  $G_{tot}$ ) and structural hierarchy (myelination) in the DK80 parcellation. The linear correlations are shown by the red line (with standard error in shaded gray) overlaid on the scatterplots (see Results for  $r$  values). The results in this coarser-scale DK80 parcellation are fully consistent with the finer-scale Glasser parcellation.

in tasks and rest for  $G_{in}$  (receivers), which reflects that different tasks process incoming information differently. On the other hand, outgoing flow of information (drivers) is similar across tasks, suggesting that sensory areas are consistently driving the information flow. Importantly, as shown in the  $G_{tot}$  matrix, the total processing of information flow is more similar within the seven tasks than compared with rest.

**Quantifying the FRIC in tasks and resting state.** In the structural domain, Zamora-Lopez, Van Heuvel and Sporns proposed the concept of a structural ‘rich club’<sup>5,7</sup>, which is characterized by a tendency for high-degree brain regions to be more densely connected among themselves than regions of a lower degree, providing important information on the higher-level topology of the brain network. We extend this concept to the functional domain by defining the concept of a FRIC, which crucially is not static but changes between tasks and rest (see Methods).

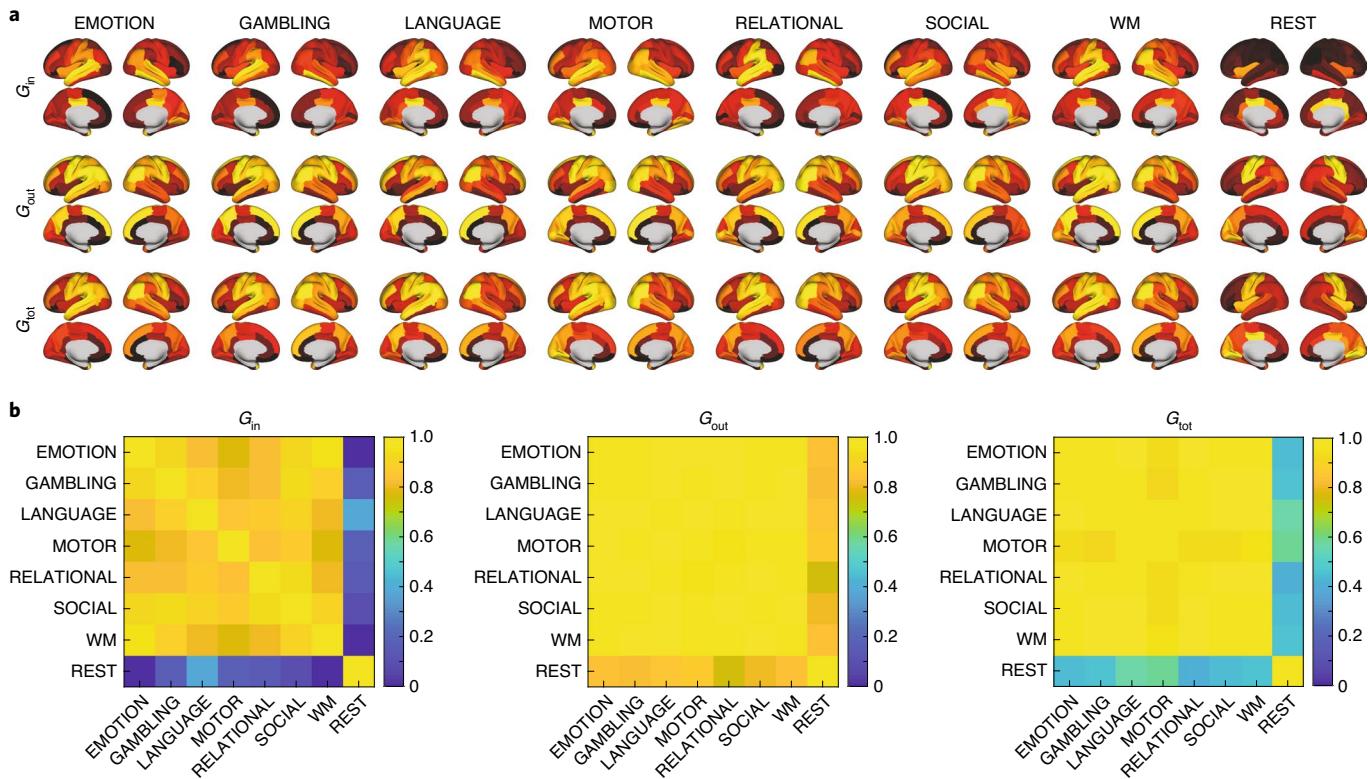
We computed the NDTE matrices for the DK80 parcellation for all seven tasks and resting state. This allows us to compute the FRIC as the set of regions that define a ‘club’ of functional hubs characterized by a tendency to be more densely functionally connected among themselves than to other brain regions from where they receive integrative information (see Methods). Figure 4a shows that the FRICs across tasks and rest contain similar but not identical core regions.

**Quantifying the GW.** We quantified the GW, which we defined as the intersection of FRIC members across all possible tasks and resting state. Figure 4b plots a rendering of the cortical and subcortical regions in the GW.

We found that the intersecting FRIC members for all seven tasks and rest include the following eight brain regions: left precuneus, left nucleus accumbens, left putamen, left posterior cingulate cortex, right hippocampus, right amygdala and left and right isthmus cingulate. Searching for a less restrictive definition, we further lowered the threshold to include areas only common to seven FRICs, which adds two further regions: right nucleus accumbens and right posterior cingulate, while lowering to six FRICs adds left and right rostral anterior cingulate. Moreover, we found that lowering the threshold to five FRICs added the left amygdala and left globus pallidus internus, while lowering to four FRICs added the left parahippocampal cortex. The results point to a stable core of brain regions necessary in the GW.

In addition to computing the GW using our formal definition, we were also prompted by a peer reviewer to investigate how this might look in a finer parcellation, the Brainnetome246 (see Methods). Unfortunately, for reasons of computational complexity, it is not possible to use our formal definition on this finer parcellation. We therefore built a method to approximate the GW as the intersection of the regions with the highest incoming information flow across tasks and rest (see Methods). Extended Data Fig. 5 shows an overlap between this rough approximation in Brainnetome246 and the fully computed GW (as the intersection of FRICs in tasks and rest) in DK80.

**Comparison with the anatomical rich club.** The proposed FRIC concept is functional in nature, thus changing with task and capturing the top regions involved in the relevant functional organization of the task. This is different from the existing anatomical rich club concept<sup>5,7</sup>, which is static in nature. Extended Data Fig. 6 shows the



**Fig. 3 | Functional hierarchy across different tasks and rest.** **a**, Cortical renderings of all seven tasks and rest for the incoming ( $G_{in}$ ), outgoing ( $G_{out}$ ) and total ( $G_{tot}$ ) directional information flow computed from the NDTE matrix of 1,003 HCP participants (see Fig. 1 and Methods) in the DK80 parcellation (using 3D views from the side and midline). **b**, Matrices of the comparison of incoming ( $G_{in}$ ), outgoing ( $G_{out}$ ) and total ( $G_{tot}$ ) directional information flows in the seven tasks and rest. As can be clearly seen, the  $G_{in}$  matrix and the renderings of the incoming flows of information (receivers) are significantly different between tasks and rest. This suggests that different tasks process incoming flows differently. This is in contrast to the  $G_{out}$  matrix and the renderings of the outgoing flows of information (drivers), which are similar. This shows that sensory areas are consistently driving the information flow. Interestingly, as can be seen from the  $G_{tot}$  matrix, the total processing of information flow is more similar within the seven tasks than compared with rest, suggesting the extrinsic, sensory nature of task processing compared with the intrinsic nature of resting-state processing.

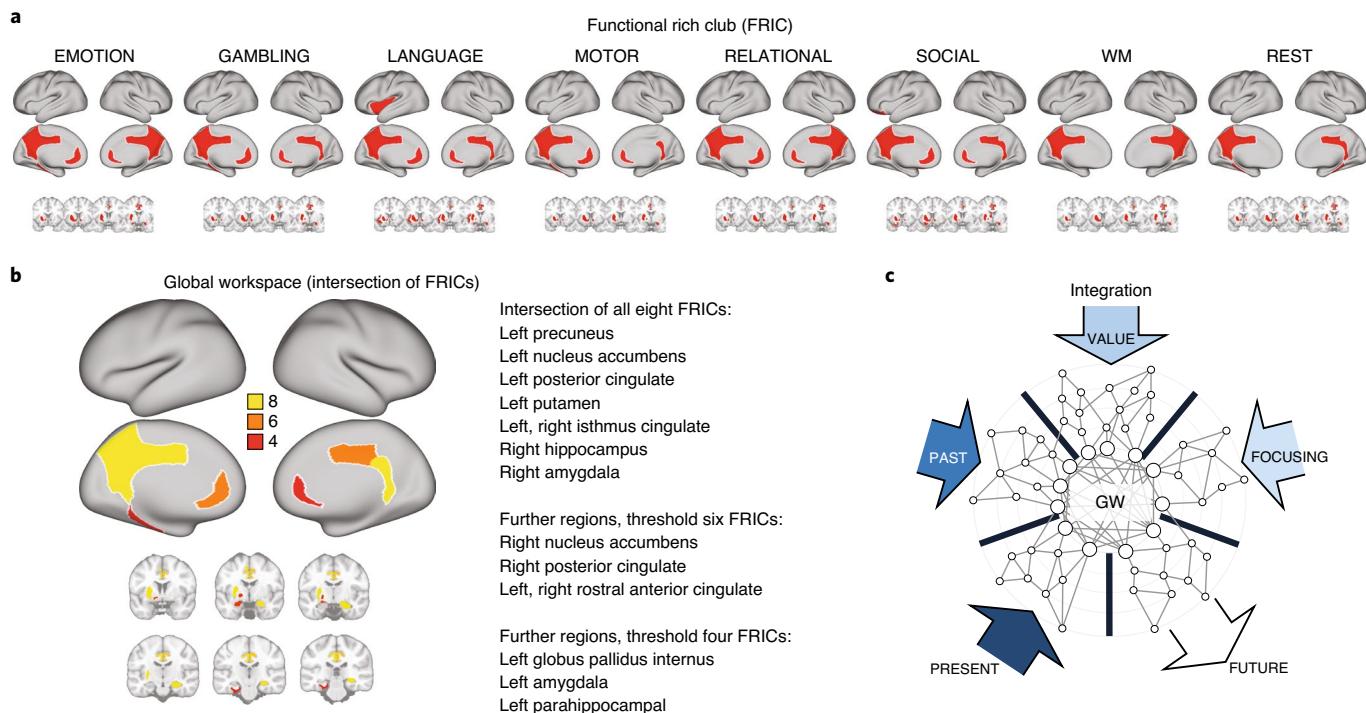
result of computing the anatomical rich club for the structural DK80 connectivity matrix and the comparison with the FRICs of the seven tasks and their intersection in the GW. The results show that there are ten brain regions in the anatomical rich club: left and right precuneus, left and right thalamus, left and right superior frontal, left superior parietal, left insula and right isthmus cingulate cortices. Only two regions are shared with the GW, namely the left precuneus and the right isthmus cingulate cortex. Equally, the FRICs of the individual tasks and rest show a similar overlap. This demonstrates the relevance of both concepts but highlights the dynamic nature of the FRIC concept, enabling spatiotemporal characterization of the functional hierarchical organization of the brain.

**Establishing the mechanistic significance of FRIC by lesioning a whole-brain model.** To establish the mechanistic functional relevance of the FRIC brain regions orchestrating the functional hierarchical organization of the resting state, we derived a causal quantification from simulation. We created a whole-brain model that can generate the effective connectivity underlying the functional hierarchy (see Methods). Figure 5a shows the general framework of how this model uses optimized anatomically constrained parameters (generative anatomically constrained bidirectional connectivity (GABIC)) to describe the effective strength of the synaptic coupling to fit the NDTE matrix by maximizing the fit of the empirical and simulated NDTE matrices. The NDTE matrix is a measure of the effective connectivity in the brain, and consequently

GABIC is a matrix generating this effective connectivity rather than the effective connectivity per se. The Hopf model was constructed using both the standard and special structural connectivity (SC) matrices (see Methods), where the standard matrix is estimated on 985 HCP participants while the special SC matrix is estimated using a higher-quality diffusion magnetic resonance imaging (dMRI) protocol but on only 32 HCP participants. Extended Data Fig. 7 shows that they are highly correlated ( $r(3,160)=0.85$ ,  $P<0.001$ , 95% CI 0.84–0.86) but that there are slightly more interhemispheric connections in the special SC matrix compared with the standard SC matrix, suggestive of a better estimation of the true SC.

Optimizing the Hopf models using a particle swarm optimizer (see Methods) was computationally very demanding, ultimately yielding a fit with a correlation between empirical and model-generated NDTE matrices of 0.6 for both the standard and special SC matrix. Furthermore, the model NDTEs produced with both SC matrices were very highly correlated. Note that the NDTE matrix is bidirectional, thus the GABIC matrix is also asymmetric. We found no statistically significant evidence that the GABIC matrix correlated with the NDTE matrix (correlation  $r(3,160)=0.0036$ , not significant, 95% CI 0.002–0.0047). NDTE correlated  $r(3,160)=0.51$ ,  $P<0.001$ , 95% CI 0.511–0.513) with static functional connectivity, demonstrating that there is complementary information in the NDTE, thus further constraining the whole-brain modelling.

Figure 5b shows the result of lesioning all FRIC compared with non-FRIC members in the whole-brain model, which causes a



**Fig. 4 | Identifying the GW as the intersection of FRICs for rest and seven tasks.** **a**, We computed the functional hierarchical organization of all seven tasks (emotion, gambling, language, motor, relational, social and working memory) and rest for the HCP participants. This allows us to compute the FRIC as the set of regions that define a 'club' of functional hubs characterized by a tendency to be more densely functionally connected among themselves than to other brain regions from where they receive integrative information (see Methods). As can be seen, these FRICs are similar but not identical across tasks and rest. **b**, We compute the regions in the GW as the intersection of the FRIC members across all possible tasks and resting state. Here, we used the maximal amount of tasks available to provide a reliable estimate of the GW. At the bottom of the figure, we show a rendering of the cortical and subcortical regions in the GW. The FRIC regions for all seven tasks and rest defining the GW are the following eight brain regions: left precuneus, left nucleus accumbens, left putamen, left posterior cingulate cortex, right hippocampus, right amygdala and left and right isthmus cingulate. Lowering the threshold for participation to more than six FRICs adds two further regions: right nucleus accumbens and right posterior cingulate (in seven FRICs) and left and right rostral anterior cingulate (in six FRICs). Further lowering the threshold to four FRICs includes another three brain regions: left amygdala and left globus pallidus internus (in five FRICs) and left parahippocampal cortex (in four FRICs). **c**, These regions fit well with the idea suggested by Dehaene and Changeux that the GW is ideally placed for integrating information from perceptual (PRESENT), long-term memory (PAST), evaluative (VALUE) and attentional (FOCUSING) systems.

significant breakdown in information flow. When measuring this in terms of the correlation between the empirical and simulated NDTE matrices, the relevant comparisons revealed a highly significant change in the fitting of the model when FRIC members are lesioned compared with the lesioning of non-FRIC members selected from all other regions. Specifically, we ran Wilcoxon rank-sum comparisons of the different conditions. First, we found that the correlation between the empirical (Emp) and simulated NDTE matrices is maximally affected when FRIC members are lesioned ( $P < 0.001$ , median difference: 0.28, 95% CI 0.27–0.29, compare first and second bars in Fig. 5b) compared both with non-FRIC members that were randomly selected (nFRIC) ( $P < 0.001$ , median difference 0.16, 95% CI 0.13–0.19, compare first and third bars in Fig. 5b) and with non-FRIC members that were selected using a biased method where we chose the regions with the top 50%  $G_{in}$  values (nFRIC<sub>bias</sub>) ( $P < 0.001$ , median difference 0.16, 95% CI 0.15–0.18, compare first and fourth bars in Fig. 5b). We also compared FRIC versus nFRIC ( $P < 0.001$ , median difference 0.12, 95% CI 0.07–0.16, compare second and third bars in Fig. 5b) and FRIC versus nFRIC<sub>bias</sub> ( $P < 0.001$ , median difference 0.12, 95% CI 0.08–0.15, compare second and fourth bars in Fig. 5b).

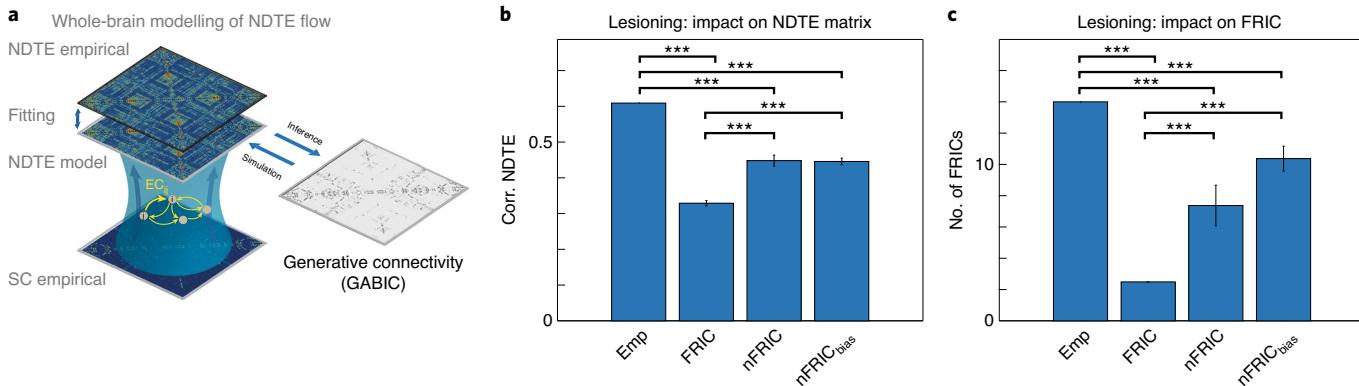
Similarly, we found significant differences in the number of FRIC members detected when lesioning the original empirical FRIC members compared with empirical non-FRIC members ( $P < 0.001$ ,

median difference 11.51, 95% CI 11.50–11.52, compare first and second bars in Fig. 5c), Emp versus nFRIC ( $P < 0.001$ , median difference 6.63, 95% CI 4.03–9.23, compare first and third bars in Fig. 5c) and Emp versus nFRIC<sub>bias</sub> ( $P < 0.001$ , median difference 3.63, 95% CI 2.04–5.22, compare first and fourth bars in Fig. 5c). We also compared FRIC versus nFRIC ( $P < 0.001$ , median difference 4.88, 95% CI 2.27–7.49), compare second and third bars in Fig. 5c and FRIC versus nFRIC<sub>bias</sub> ( $P < 0.001$ , median difference 7.88, 95% CI 6.27–9.49, compare second and fourth bars in Fig. 5c). This provides mechanistic evidence for the role of FRIC in orchestrating functional organization.

**Alternative measurement of hierarchy.** In addition to the NDTE measures of connection strength, Extended Data Fig. 8 shows the result of measuring another hierarchical measurement inspired by the seminal generalized linear model (GLM) method for computing the fraction of feedforward and feedback (FF) hierarchical organization, which was originally developed in neuroanatomy<sup>1,3,41</sup> (see Methods).

## Discussion

The acquisition of large multimodal neuroimaging datasets (Big Data) is now making it possible to address fundamental problems in systems neuroscience<sup>18,42,43</sup>. A key open problem is how best to



**Fig. 5 | Establishing causal significance by lesioning whole-brain model.** **a**, To establish the causal mechanistic functional relevance of the FRIC brain regions orchestrating the functional hierarchical organization of the resting state, we created a whole-brain model to fit the empirical NDTE flow matrix. **b**, Lesioning all FRIC compared with non-FRIC members in this whole-brain model causes a significant breakdown in information flow (see Results for significance testing). The correlation between empirical (Emp) and simulated NDTE matrices is maximally affected when FRIC members are lesioned compared both with non-FRIC members that were randomly selected (nFRIC) and with non-FRIC members that were selected using a biased method in which we chose the regions with the top 50%  $G_{in}$  values (nFRIC<sub>bias</sub>). We also compared FRIC versus nFRIC and FRIC versus nFRIC<sub>bias</sub>. **c**, Similarly, we found significant differences in the number of FRIC members detected when lesioning the original empirical FRIC members compared with empirical non-FRIC members, Emp versus nFRIC and Emp versus nFRIC<sub>bias</sub>. We also compared FRIC versus nFRIC and FRIC versus nFRIC<sub>bias</sub>. This is mechanistic evidence for the significance of FRIC in orchestrating functional organization. Emp, empirical; FRIC, functional rich club; nFRIC, nFRIC<sub>bias</sub>, non-functional rich club members.

characterize the functional hierarchical organization of whole-brain dynamics to understand the orchestration of brain processing. We quantified the ‘GW’<sup>14,15</sup> as the core common FRIC regions invariant across seven tasks and rest in 1,003 participants<sup>19</sup>. To estimate the overall functional hierarchical organization of the human brain and the corresponding regions in the FRIC in resting state and across seven tasks, we developed an NDTE framework, validated through retesting (Extended Data Fig. 3) and external validation (see Extended Data Fig. 9 and Methods). This provided the possibility to precisely characterize functional hierarchical organization by estimating the flow of information between pairs of brain regions. Finally, the importance of the findings was confirmed by building and lesioning a whole-brain model, which not only gave a mechanistic understanding of the emergence of the FRIC but also showed that the regions in the FRIC are orchestrating brain function.

The GW was found to consist of a core subset of brain regions including the precuneus, posterior and isthmus cingulate, nucleus accumbens, putamen, hippocampus and amygdala. This core functional ‘club’ of integrative brain regions is consistent with the original proposal by Dehaene and Changeux<sup>15</sup>, which suggests that the global neuronal workspace must integrate past and present through focusing and evaluation. Indeed, Dehaene and Changeux proposed that associative perceptual, motor, attention, memory and value areas interconnect to form a higher-level unified space. For the integration of the past, the hippocampus has been shown to play a key role in many aspects of memory (see, for example, refs. 44–46). Similarly, the evaluation of value has been shown to involve the nucleus accumbens (see, for example, refs. 47–49), putamen (see, for example, refs. 49,50) and amygdala (see, for example, refs. 49,51–54). The integration of the past, present and future by processing and attending perceptual information has been strongly associated with the precuneus (see, for example, refs. 55–57) and the posterior and isthmus cingulate cortices (see, for example, refs. 49,57–60). Interestingly, the functions of the precuneus have also been shown to be compromised in coma and vegetative state<sup>61</sup>.

Baars’ cognitive theory of consciousness distinguishes a vast array of unconscious specialized processors running in parallel, and a single limited-capacity serial GW that allows them to exchange information<sup>14</sup>. The subsequent development by Dehaene

and Changeux<sup>62</sup> of this theory into the global neuronal workspace includes a further ‘ignition’ component capturing the strong temporary increase in synchronized firing leading to a coherent state of activity. The transition to this state of highly correlated activity is very fast and leads to amplification of local neural activation and the subsequent ignition of multiple distant areas. We have not studied ignition here since the fast timescale (typically about 100–200 ms) is difficult to capture with fMRI—although we have recently shown that such fast timescales can be reconstructed using appropriate whole-brain modelling<sup>63</sup>. Yet, the potential role of ignition in initiating and sustaining FRIC and the GW should clearly be further investigated in future studies using, for example, MEG.

The identification of the GW using whole-brain modelling of the established information flow (with NDTE) in the empirical data complements existing methods for establishing information<sup>20–22</sup> and functional gradiental hierarchy<sup>8,64</sup>.

Our findings are consistent with the gradiental perspective on hierarchical processing<sup>8,64</sup> but goes beyond these findings due to the ability of our NDTE framework to estimate bidirectional information flow in both resting and tasks. We note that, even though we have used the most comprehensive dataset currently available with rest and tasks covering a broad cognitive domain, in the future it would be of considerable interest to compare the results in an even larger set of tasks and participants.

A long history of neuroanatomical discoveries has demonstrated that the brain is clearly hierarchical in its structure, from single units to the larger circuits<sup>1–7</sup>. Research by Margulies and colleagues<sup>8</sup> used neuroimaging to extend Mesulam’s seminal proposal that brain processing is shaped by a hierarchy of distinct unimodal areas to integrative transmodal areas<sup>2</sup>. Recently, this idea has been further extended by applying the principle of harmonic modes to functional connectivity HCP data<sup>64</sup>.

We extend these findings using the advanced NDTE framework to establish bidirectional information flow for identifying the functional hierarchical organization of the human brain. We have demonstrated that this can be used to characterize the FRIC corresponding to the core integrative transmodal brain regions allowing for the necessary whole-brain cohesion (as shown by the cartoon in a simplified 2D representation in Fig. 1c). As shown in

Fig. 2, the outgoing information flow (drivers),  $G_{\text{out}}$ , reflects mostly the unimodal sensory regions, whereas the incoming information flow (integrative receivers),  $G_{\text{in}}$ , reflects the higher-order transmodal regions.

As shown by our results here, the NDTE framework can be used for any spatial scale and yields very consistent results from the three parcellations we have used here (DK80, Brainnetome246 and Glasser360). Still, it is important to note that there is no current consensus about what is the correct spatial parcellation scheme, as shown by a recent paper by Eickhoff and colleagues reviewing the literature on the topographic organization of the brain<sup>65</sup>. Nevertheless, the results are similar and consistent at the macroscale, reflecting the functional organization of the human brain at different spatial scales.

Furthermore, we also confirmed the results by applying the NDTE framework to MEG time series from the 62 cortical regions of the DK80 parcellation, and demonstrating similarly high correlations with myelination (Extended Data Fig. 2). This clearly demonstrates the robustness of the NDTE framework for both haemodynamic and direct electromagnetic measures of brain activity.

The NDTE framework includes suitable normalization, surrogates and  $P$  value aggregation across large number of participants (see Methods), which contribute to the robustness of the method. Indeed, the high correlation between myelination and the total incoming and outgoing information flow,  $G_{\text{tot}}$ , obtained for both blood-oxygen-level-dependent (BOLD) and MEG signals, provides further confidence in the results.

To provide further evidence of the causal significance of the core FRIC regions, we built and lesioned a whole-brain model that can generate effective connectivity obtained using our NDTE framework (see Fig. 5 and Methods). We found that lesioning ten FRIC regions for the resting state very significantly impaired the flow of information and the ability to form new FRIC among the remaining brain regions. This causally establishes the full FRIC network as having a key integrative role in orchestrating functional hierarchical organization. Furthermore, we found that this FRIC network is fairly fault tolerant in that lesioning four FRIC members was not significantly different from lesioning four non-FRIC members. However, lesioning five FRIC compared with five non-FRIC members did lead to significantly different information flow. This suggests that there is a tipping point beyond which the breakdown of information flow in the FRIC leads to important problems. It is possible that the partial breakdown of FRIC members could be a considerable factor in the transitioning to neuropsychiatric disorders<sup>66</sup>. In particular, this leads to the speculation that the specific breakdown of the integration of the evaluative system related to self-processing (precuneus, nucleus accumbens, putamen and cingulate cortices) could lead to the anhedonia, the lack of pleasure, which is a major symptom of neuropsychiatric disease<sup>67,68</sup>.

Traditionally, whole-brain models have been relatively successful in linking SC with functional dynamics<sup>66,69,70</sup>. This has revealed important mechanistic principles of brain function<sup>63,71–74</sup>. In particular, the Hopf model has been shown to be useful for modelling both fMRI<sup>73,75–77</sup> and MEG functional connectivity data<sup>78</sup>. In fact, we chose the Stuart-Landau Hopf model, since it has been shown that this is the most universal non-linear oscillator, being applied broadly across a large set of completely different physical systems<sup>79</sup>.

The findings presented here help shed light on a major unsolved problem in neuroscience. While the results presented here pertain to the GW of conscious processing, future work could use the NDTE framework to investigate other states such as sleep and anaesthesia, allowing for direct comparison with other theories of consciousness<sup>80,81</sup>. Equally, the NDTE framework could be used to investigate unbalanced brain states in neuropsychiatric disorders and be used to perturb and rebalance the model to identify novel optimal, causal paths to health<sup>66,73,82</sup>.

## Methods

### Neuroimaging data acquisition, preprocessing and time series extraction.

**Ethics.** The Washington University–University of Minnesota (WU-Minn HCP) Consortium obtained full informed consent from all participants, and research procedures and ethical guidelines were followed in accordance with Washington University institutional review board approval (Mapping the Human Connectome: Structure, Function and Heritability; IRB # 201204036).

**Participants.** The dataset used for this investigation was selected from the March 2017 public data release from the Human Connectome Project (HCP), where we chose a sample of 1,003 participants, all of whom having resting-state data. For the seven tasks, HCP provides the following numbers of participants: working memory (WM), 999; SOCIAL, 996; MOTOR, 996; LANGUAGE, 997; GAMBLING, 1,000; EMOTION, 992; RELATIONAL, 989. We also validated the framework in the 45 participants with retest data. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous publications using the full HCP dataset.

**Neuroimaging acquisition for fMRI HCP.** The 1,003 HCP participants were scanned on a 3-T connectome-Skyra scanner (Siemens). We used one resting-state fMRI acquisition of approximately 15 min acquired on the same day, with eyes open with relaxed fixation on a projected bright cross-hair on a dark background as well as data from the seven tasks. The HCP website (<http://www.humanconnectome.org/>) provides the full details of participants, the acquisition protocol and preprocessing of the data for both resting state and the seven tasks. Below we have briefly summarized these.

**Neuroimaging acquisition for dMRI HCP.** We obtained multi-shell diffusion-weighted imaging data from 985 subjects of the HCP 1200 data release. The standard acquisition protocol takes 59 min (six runs, each of approximately 9 min 50 s). We also obtained diffusion spectrum and T2-weighted imaging data from 32 participants from the HCP database who were scanned for a full 89 min. The acquisition parameters for both groups are described in detail on the HCP website<sup>83</sup>.

**Neuroimaging acquisition for MEG HCP.** We used the human non-invasive resting-state MEG data publicly available from the Human Connectome Project (HCP) consortium, acquired on a Magnes 3600 MEG (4D NeuroImaging) with 248 magnetometers. The resting-state data consist of 89 subjects (mean 28.7 years, range 22–35 years, 41 female/48 male, acquired in three subsequent sessions lasting 6 min each).

**The HCP task battery of seven tasks.** The HCP task battery consists of seven tasks: working memory, motor, gambling, language, social, emotional and relational, which are described in detail on the HCP website<sup>40</sup>. HCP states that the tasks were designed to cover a broad range of human cognitive abilities in seven major domains that sample the diversity of neural systems: (1) visual, motion, somatosensory and motor systems; (2) working memory, decision-making and cognitive control systems; (3) category-specific representations; (4) language processing; (5) relational processing; (6) social cognition; (7) emotion processing. In addition to resting-state scans, all 1,003 HCP participants performed all tasks in two separate sessions (first session: working memory, gambling and motor; second session: language, social cognition, relational processing and emotion processing). As a test-retest control condition, a small subsample of 45 HCP participants performed the paradigm twice.

**Neuroimaging SC and extraction of functional time series. Parcellations.** All neuroimaging data were processed using two standard cortical parcellations with added subcortical regions. For a fine-scale parcellation, we used the Glasser parcellation with 360 cortical regions (180 regions in each hemisphere)<sup>28</sup>. We added the 18 subcortical regions, that is 9 regions per hemisphere: hippocampus, amygdala, subthalamic nucleus (STN), globus pallidus internal segment (GPi), globus pallidus external segment (GPe), putamen, caudate, nucleus accumbens and thalamus. This created a parcellation with 378 regions, that is, the Glasser378 parcellation, which is defined in the common HCP Connectivity Informatics Technology Initiative (CIFTI) grayordinates standard space with a total of 91,282 grayordinates (sampled at 2 mm<sup>3</sup>).

For a coarser-scale parcellation, we used the Mindboggle-modified Desikan-Killiany parcellation<sup>30</sup> with a total of 62 cortical regions (31 regions per hemisphere)<sup>31</sup>. We added the same 18 subcortical regions mentioned above (9 regions per hemisphere) and ended up with 80 regions in the DK80 parcellation, also precisely defined in the common HCP CIFTI grayordinates standard space.

For a medium-scale parcellation, we used the Brainnetome parcellation with a total of 246 regions including 210 cortical regions (105 regions per hemisphere) and 36 subcortical regions (18 per hemisphere)<sup>29</sup>. Finally, to compare the incoming and outgoing information flow with existing networks, we used the very coarse-scale Yeo parcellation with seven regions<sup>84</sup> (see Supplementary Fig. 1).

**Generating SC matrices from dMRI.** To be as precise as possible for the model fitting, we estimated the SC matrix from two HCP dMRI datasets. The first dataset,

standard HCP dMRI (<http://www.humanconnectome.org/>), uses the highest quality multi-shell diffusion data acquired in sequence, taking 59 min from 985 HCP participants (HCP data acquired at Washington University in St. Louis; see Acknowledgements<sup>19,85</sup> (for HCP specifications, see their website). The second dataset, special HCP dMRI (<http://www.humanconnectome.org/>), uses even better protocols, taking 89 min for each of 32 HCP participants at the Massachusetts General Hospital centre. Both dMRI datasets were preprocessed and made available as part of the freely available Lead-DBS software package (<http://www.lead-dbs.org/>).

The precise preprocessing is described in detail in Horn and colleagues<sup>86</sup>, but briefly, the data were processed using a generalized  $q$ -sampling imaging algorithm implemented in DSI Studio (<http://dsi-studio.labsolver.org>). Segmentation of the T2-weighted anatomical images produced a white-matter mask and co-registering of the images to the b0 image of the diffusion data using SPM12. In each HCP participant, 200,000 fibres were sampled within the white-matter mask. Fibres were transformed into Montreal Neurological Institute (MNI) space using Lead-DBS<sup>87</sup>. The methods used the algorithms for false-positive fibres shown to be optimal in recent open challenges<sup>88,89</sup>. The risk of false-positive tractography was reduced in several ways. Most importantly, this used the tracking method achieving the highest (92%) valid connection score among 96 methods submitted from 20 different research groups in a recent open competition<sup>88</sup>. We subsequently used the standardized methods in Lead-DBS to produce the structural connectomes for the DK80 parcellation used in the whole-brain Hopf model (see section "Whole-brain model of NDTE flow").

**Preprocessing and extraction of functional time series in fMRI resting state and task data.** The preprocessing of the HCP resting state and task datasets is described in detail on the HCP website. Briefly, the data are preprocessed using the HCP pipeline, which is based on standardized methods using FSL (FMRIB Software Library), FreeSurfer and the Connectome Workbench software<sup>90,91</sup>. This standard preprocessing included correction for spatial and gradient distortions and head motion, intensity normalization and bias field removal, registration to the T1-weighted structural image, transformation to the 2-mm MNI space, and using the FIX artefact removal procedure<sup>91,92</sup>. The head motion parameters were regressed out and structured artefacts were removed by independent component analysis followed by FMRIB's ICA-based X-noiseifier (ICA+FIX) processing<sup>93,94</sup>. Preprocessed time series of all grayordinates are in HCP CIFTI grayordinates standard space and available in the surface-based CIFTI file for each participants for resting state and each of the seven tasks.

We used a custom-made MATLAB script using the 'ft\_read\_cifti' function (Fieldtrip toolbox<sup>95</sup>) to extract the average time series of all the grayordinates in each region of the Glasser378, Brainnetome246 and DK80 parcellations, which are defined in the HCP CIFTI grayordinates standard space. The BOLD time series were filtered using a second-order Butterworth filter in the range of 0.008–0.08 Hz.

**Preprocessing and extraction of MEG data time series.** For each participant, the MEG data were acquired in a single continuous run comprising resting state. As a starting point, we used the preprocessed MEG data from the HCP database. At this level of preprocessing, removal of artefactual independent components, bad samples and channels has already been performed. Following the preprocessing pipeline detailed in ref.<sup>96</sup>, MEG data were then downsampled to 250 Hz using an anti-aliasing filter, filtered to remove frequencies below 1 Hz, co-registered with the head models provided by the HCP and source-reconstructed using linearly constrained minimum-variance beamforming<sup>97</sup> to the 62 cortical regions of the DK80 parcellation. For each region, a single time series was computed as the first principal component of the voxels within that parcel. Each region's time series was then standardized to have mean of zero and standard deviation of one, so that the amount of variance was always the same regardless of the depth of the region. To project the results to brain space, we used a weighted mask, where each region had its maximum value at the centre of gravity.

We did not apply correction for spatial leakage (volume conduction), because the amount of signal removal would have been too large given the number of regions in the parcellation, thus making the rest of the analysis impossible. Although the similarity with the fMRI results suggests that the impact of signal leakage is not too disruptive for the conclusions of this study, this remains a limitation of this work, and future efforts will be dedicated to rigorous assessment of the impact of different approaches to leakage correction on the NDTE metric, as well as to the consideration of 'softer' versions of such leakage correction approaches that can be used in this case.

**Neuroimaging analysis tools and methods. Normalized directed transfer entropy.** To establish and investigate the functional hierarchical organization of whole-brain activity, we first need to characterize how different brain regions communicate with each other, that is, to compute the directed flow between regions. We characterize the functional interaction between two brain regions, in a given parcellation, by an information-theoretical statistical criterion that allows us to infer the underlying bidirectional reciprocal communication. The NDTE framework was inspired by the work of Brovelli and colleagues, who used and validated a similar transfer entropy framework for neuroimaging data<sup>25</sup>. This framework uses a Gaussian

approximation, that is, only second-order statistics of the involved entropies, which means, as shown below, that instead of estimating the probabilities, the method estimates the covariance, which massively facilitates computation. Finally, as also outlined below, we add four key elements to this powerful framework: normalization, multiple timepoints in the past, circular surrogates and aggregation of  $P$  values to improve the reliability and robustness of the NDTE framework.

Let us assume that we want to describe the statistical causal interaction exerted from a source brain area  $X$  to another target brain area  $Y$ . We aim to measure the extra knowledge that the dynamical functional activity of the past of  $X$  contributes to the prediction of the future of  $Y$ , by the following mutual information:

$$I(Y_{i+1}; X^i | Y^i) = H(Y_{i+1} | Y^i) - H(Y_{i+1} | X^i, Y^i) \quad (1)$$

where  $Y_{i+1}$  is the activity level of brain area  $Y$  at time point  $i + 1$ , and  $X^i$  indicates the whole activity level of the past of  $X$  (filtered BOLD signal sampled in repetition time (TR) in a time window of length  $T$  up to and including the time point  $i$  (that is,  $X^i = [X_i, X_{i-1}, \dots, X_{i-(T-1)}]$ ). Following Kantz and Schreiber<sup>98</sup>, we estimated the autocorrelations of the empirical BOLD signals for all regions and all participants, finding that on average this decay to the first minimum corresponded to a value of  $T = 10$ . Note that this causality measure is not symmetric, which allows bidirectional analysis. The conditional entropies are defined as follows:

$$\begin{aligned} H(Y_{i+1} | Y^i) &= H(Y_{i+1}, Y^i) - H(Y^i) \\ &= - \sum_{y_{i+1}, y^i} p(y_{i+1}, y^i) \log(p(y_{i+1} | y^i)) \end{aligned} \quad (2)$$

$$\begin{aligned} H(Y_{i+1} | X^i, Y^i) &= H(Y_{i+1}, Y^i, X^i) - H(X^i, Y^i) \\ &= - \sum_{y_{i+1}, y^i, x^i} p(y_{i+1}, y^i, x^i) \log(p(y_{i+1} | x^i, y^i)) \end{aligned} \quad (3)$$

The mutual information  $I(Y_{i+1}; X^i | Y^i)$  expresses the degree of statistical dependence between the past of  $X$  and the future of  $Y$ . In other words, if that mutual information is equal to zero, then the probability  $p(Y_{i+1}, X^i | Y^i) = p(Y_{i+1} | Y^i) p(X^i | Y^i)$ , and thus we can say that there is no causal interaction from  $X$  to  $Y$ .

Consequently  $I(Y_{i+1}; X^i | Y^i)$  expresses a strong form of Granger causality<sup>99</sup>, by comparing the uncertainty in  $Y_{i+1}$  when using knowledge of only its own past  $Y^i$  or the past of both brain regions, that is,  $X^i, Y^i$ . This information-theoretical concept of causality was introduced in neuroscience by Schreiber<sup>100</sup> and is usually called the transfer entropy<sup>25,101–103</sup>. To facilitate computation, Brovelli et al.<sup>25</sup> proposed a weaker form of causality allowing the calculation of the involved entropies by just considering a Gaussian approximation, that is, by considering only second-order statistics. Indeed, under this approximation, the entropies can be computed as follows:

$$H(Y^i) = \frac{T}{2} \log(2\pi e) + \frac{1}{2} \log(\det(\Sigma(Y^i))) \quad (4)$$

$$H(Y_{i+1}, Y^i) = \frac{T+1}{2} \log(2\pi e) + \frac{1}{2} \log(\det(\Sigma(Y_{i+1}, Y^i))) \quad (5)$$

$$H(X^i, Y^i) = T \log(2\pi e) + \frac{1}{2} \log(\det(\Sigma(X^i, Y^i))) \quad (6)$$

$$H(Y_{i+1}, Y^i, X^i) = \frac{2T+1}{2} \log(2\pi e) + \frac{1}{2} \log(\det(\Sigma(Y_{i+1}, Y^i, X^i))) \quad (7)$$

In other words, causality is based only on the corresponding covariance matrices.

To be able to sum and compare the directed mutual information flow between different pairs of brain regions, this has to be appropriately normalized. In fact, if the mutual information directed flow is correctly normalized, then the different values could be combined, for example, to determine the total directed flow exerted by the whole brain on a single region or, vice versa, the directed flow exerted by a single brain region on the whole brain. More specifically, we define this information-theoretical measure as the NDTE flow  $F_{XY}$  from time series  $X$  to  $Y$ :

$$F_{XY} = I(Y_{i+1}; X^i | Y^i) / I(Y_{i+1}; X^i, Y^i) \quad (8)$$

where  $I(Y_{i+1}; X^i, Y^i)$  is the mutual information that the past of both signals together,  $X^i, Y^i$ , has about the future of the target brain region  $Y_{i+1}$ . Given that

$$I(Y_{i+1}; X^i, Y^i) = I(Y_{i+1}; Y^i) + I(Y_{i+1}; X^i | Y^i) \quad (9)$$

this normalization compares the original mutual information directed flow, that is, the predictability of  $Y_{i+1}$  by the past of  $X^i | Y^i$ , with the internal predictability of  $Y_{i+1}$ , that is  $I(Y_{i+1}; Y^i)$ .

We define the matrix of NDTE flow in the brain for a given parcellation with  $N$  regions,  $C$ , with elements  $C_{ij} = F_{X(i)X(j)}$ , where  $X(k)$  corresponds to the filtered BOLD time series from region  $k$ .

We note that it has been shown that transfer entropy can be modulated and influenced by the frequencies and bandwidth of the filter of high-temporal-resolution signals coming from MEG/electroencephalography<sup>20,104</sup>, but these potential obstacles related to the high temporal resolution are insignificant for the slow, narrowly filtered (0.008–0.08 Hz) BOLD signal. We show that the NDTE DK80 matrices for the filtered and non-filtered version of resting state for 100 unrelated HCP participants show a correlation of 0.97 (Extended Data Fig. 10).

Furthermore, to perform a statistical significance analysis of the NDTE flow,  $C$ , we use the surrogate framework, inspired by the work of Theiler and colleagues<sup>105</sup>. This traditional surrogate methodology uses a phase randomization of the Fourier transform of the original data to preserve the linear correlations. Nevertheless, as discussed and analysed rigorously in Diks and Fang<sup>106</sup>, these methods are not suitable for detecting significance when using entropy measures, as discussed extensively<sup>107,108</sup>. In view of these problems, Quiroga and colleagues proposed the circular time-shifted surrogates method, which is a method using surrogates that can be used for causality measurements<sup>109</sup>.

There have been many suggestions of how best to construct surrogate data; see, for example, ref. <sup>109</sup>. A first approach is to construct time series with the same power spectrum (that is, an autocorrelation function) of the original ones, but with chance cross correlation. A more refined approach is to construct surrogates that also maintain the amplitude distribution (AAFT)<sup>110</sup>. However, this is based on an iterative procedure that does not always converge. In contrast, in ref. <sup>27</sup> a very simple solution was proposed: to shift one dataset with respect to the other. This way, the properties (amplitude and autocorrelation spectrum, etc.) of each signal are maintained because the signals are exactly the same, but the cross-correlation is reduced to chance levels, which is exactly what is required, as the hypothesis tested is whether a synchronization or directionality value is larger than chance. Furthermore, the use of time-shifted surrogates has already been shown to provide optimal results with several linear and non-linear synchronization measures<sup>27,109</sup>.

Hence, we use the circular shift methodology for analysing the  $P$  values of the hypothesis testing, aiming to detect significant values in  $C$  for each pair for each single participant. For each statistical test (that is, each pair of regions and each subject), we generate 100 independent circular time-shifted surrogates by separately resampling both the driving signal  $X$  and the target response signal  $Y$ . Specifically, two independent integers  $c$  and  $d$  are randomly generated within the interval [0.05n 0.95n] (where  $n$  is the number of time points in the time series signal). Then, the circular time shift is performed by moving the first  $c$  values of  $X = [X_1, \dots, X_n]$  to the end of the time series, which creates the surrogate sample  $X = [X_{c+1}, \dots, X_n, X_1, \dots, X_c]$ , and similarly for  $Y$  the first  $d$  values of  $Y = [Y_1, \dots, Y_n]$  are moved to the end of the time series to create the surrogate sample  $Y = [Y_{d+1}, \dots, Y_n, Y_1, \dots, Y_d]$ .

This type of surrogate does not assume Gaussianity and preserves the whole statistical structure of the original time series. We use a nonparametric kernel distribution representation of the probability density function of the surrogate values of  $C$ , and compare the fraction of area of that distribution above the value of the NDTE flow of the original data,  $C_{\text{original}}$ , with the total area, and compute the corresponding  $P$  value. Supplementary Fig. 2 shows all the  $P$  values for all the pairs of time series across all the participants in each of the eight conditions coming from the surrogates.

After computing the individual  $P$  values for each brain region pair and each single participant, we aggregate the  $P$  values for each single pair of brain areas across the whole group of participants. The combination of different  $P$  values across subjects is a classical problem in statistics that was originally addressed by Ronald Fisher in what is nowadays known as Fisher's method<sup>111</sup>. Here, we used a more sensitive methodology, namely the Stouffer's method<sup>112</sup>, which sums the inverse normal transformed  $P$  values. Indeed, the Stouffer's statistic is given by

$$S = \sum_i \Phi^{-1}(p_i) \quad (10)$$

where  $\Phi$  is the standard normal cumulative distribution function, and the  $p_i$  are the  $P$  values of each participant  $i$  (computed for a given pair). Under the null hypothesis, the Stouffer's statistic is normally distributed  $N(0, m)$ , where  $m$  is the total number of participants. After the aggregation of the pairs of  $P$  values across participants, we correct for multiple comparisons by using the traditional false discovery rate method of Benjamini and Hochberg<sup>112</sup>. We call the resulting NDTE matrix  $C_{\text{All}}$ .

The result of the significance test across participants determines a binary matrix  $T$  (with the dimension being the number of brain regions in a given parcellation) that indicates with ones or zeros whether the corresponding pair is significant or not (rows indicate target regions, and columns driving regions).

**Validation of NDTE framework.** First, to show that our NDTE framework captures the main causal connectivity of a given network, we use the framework and network proposed by Seth, Chorley and Barrett<sup>113</sup>. Our generative Hopf model (described in detail below) simulated BOLD signals in the proposed coupled network, which is shown graphically in Extended Data Fig. 9a and in binary matrix form in Extended Data Fig. 9b. The model used the same parameters for all the nodes as specified below in the description of the Hopf model, except using

a bifurcation parameter of  $a = -0.2$  and a global coupling strength parameter of  $G = 0.13$ . We generated time series of 1,200 points sampled with the same TR as the empirical data (0.72 s) and aggregated the run from ten repetitions. The NDTE framework was applied to the simulated data (using the same lag of 10 and 100 surrogates as in the analysis of the empirical data).

This validated the framework by recovering the causal directionality of the underlying network (compare Extended Data Fig. 9b with Extended Data Fig. 9d). We note that, for correct estimation, it is clearly important to use the three new key elements that we have added to the transfer entropy framework: normalization, circular surrogates and aggregation of  $P$  values (compare the results in Extended Data Fig. 9d with the results without these extra ingredients in Extended Data Fig. 9c).

In terms of further validation, it is important to note that there have already been important validations of the transfer entropy framework. Most importantly, it has already been shown that using surrogates is essential to avoid problems with determining directionality<sup>26</sup>. It is well known that, rather than measuring actual causality, directionality measures can reflect intrinsic properties of each of the signals. In particular, research has shown that directionality can even be reversed, which depends mainly on the noise level and the main frequencies of each of the signals. However, a solution to this problem was proposed by Quiroga<sup>27</sup>, who showed that such bias can easily be detected by using surrogates and that causality can be established using the time-shifted surrogates that we use. Another important validation is that the use of the time-shifted surrogates has been shown to give optimal results with several linear and non-linear synchronization measures<sup>27,109</sup>.

**Functional hierarchical organization.** Using the above methods describing NDTE flow, we can establish and study the functional organization of data where the different levels of directed flow to and from a given brain region  $i$  are given by  $G_{\text{in}}(i)$ ,  $G_{\text{out}}(i)$ , and  $G_{\text{tot}}(i)$ .

Our analysis of functional relevance and hierarchy is based on the resulting averaged NDTE flow,  $C_{\text{All}}$ , across all participants. We define for each brain region  $i$  the incoming level of directed flow, that is, the degree of being a receiver, by  $G_{\text{in}}(i) = \sum_j C_{\text{All}_{ij}}$ . Similarly, for each brain region  $i$  the outgoing level of NDTE flow, that is, the degree of being a driving region, is defined by  $G_{\text{out}}(i) = \sum_j C_{\text{All}_{ji}}$ . The total level of functional interaction for each brain region  $i$  is thus given by  $G_{\text{tot}}(i) = G_{\text{in}}(i) + G_{\text{out}}(i)$ .

**Functional rich club.** We define the FRIC in a matrix  $C_{\text{All}}$  with  $N$  regions by running a simple algorithm that searches for the largest subset of regions  $k = [i_1, \dots, i_l]$ , where  $G_{\text{FRIC}}(k)$  is significantly larger than all other sets with the same number of regions:

$$G_{\text{FRIC}}(k) = \sum_k C_{\text{All}_k} + \sum_k G_{\text{in}}(k) - \sum_k G_{\text{out}}(k) \quad (11)$$

Using a more detailed notation, the equation can be further expressed as  $G_{\text{FRIC}}(k) = \sum_{i=1}^l \sum_{j=1}^l C_{\text{All}_{ij}} + \sum_{i=1}^l G_{\text{in}}(i) - \sum_{j=1}^l G_{\text{out}}(j)$ , where  $l$  is the number of regions that defines the subset of regions  $k = [i_1, \dots, i_l]$ .

The first and third term can be transformed to

$$\sum_{i=1}^l \sum_{j=1}^l C_{\text{All}_{ij}} - \sum_{j=1}^l G_{\text{out}}(j) = \sum_{j=1}^l \left[ \sum_{i=1}^l C_{\text{All}_{ij}} - G_{\text{out}}(j) \right] = - \sum_{j=1}^l \sum_{i=l+1}^N C_{\text{All}_{ij}}$$

where  $N$  is the total number of regions.

Similarly, the second term, that is the sum of  $G_{\text{in}}$ , can be expressed as

$$\sum_{i=1}^l G_{\text{in}}(i) = \sum_{i=1}^l \sum_{j=l}^N C_{\text{All}_{ij}} = \sum_{j=1}^l \sum_{i=1}^l C_{\text{All}_{ij}} + \sum_{i=1}^l \sum_{j=l+1}^N C_{\text{All}_{ij}}$$

Thus, taken together, Eq. 11 can also be written as

$$G_{\text{FRIC}}(k) = \sum_{j=1}^l \left[ \sum_{i=1}^l C_{\text{All}_{ij}} - \sum_{i=l+1}^N C_{\text{All}_{ij}} \right] + \sum_{i=1}^l \sum_{j=l+1}^N C_{\text{All}_{ij}}$$

where the first term describes the difference between the information flow of each member of the subset to the other members in the subset compared with the information flow outside the subset. The second term describes the information flow that each member of the subset receives from outside.

As can be appreciated from the combinatorics, for a matrix  $C_{\text{All}}$  with more than just a few regions, it is computationally very demanding to exhaustively compute the optimal solution. However, it is also clear from the definition that the FRIC for a given set of regions is likely to be found from the regions with the highest level of incoming directed flow ( $G_{\text{in}}$ ). So we created the following algorithm: Sort the regions according to  $G_{\text{in}}$ , and then iteratively compute  $G_{\text{FRIC}}$  for progressively more  $k$  regions. Statistical significance was computed by replacing a random region of the  $k$  regions with any of the remaining regions using a Monte Carlo framework.

In more detail, the FRIC algorithm iteratively expands the FRIC by adding a new node to the existing club until the FRIC value of the FRIC club is no

longer statistically significant by comparing whether a FRIC of length  $k$  has a significantly larger FRIC value than all other possible surrogate clubs of this length. The FRIC value is computed as the sum of the connections between all club members plus the sum of all the incoming connections for the club members minus the sum of all outgoing connections for the club members, as shown in Eq. 11. The FRIC value is also computed for all the permutations of all other clubs that can be made from the same number of existing nodes (with length  $k$ ). As mentioned, given the challenges of the combinatorics, we used a Monte Carlo framework to generate the most similar and challenging permutations of the FRIC, that is, when comparing a potential new FRIC club of size  $k$ , the surrogate clubs (with  $k$  members) have the same  $k - 1$  members with an added random new member. We run this procedure iteratively, starting from the top  $G_{in}$  regions, and continue while the  $P$  value of the comparison between the FRIC club and surrogates is smaller than 0.05. We seed the procedure with the top rich club region having the largest incoming NDTE flow.

As can be appreciated from the above description, the computational complexity of the FRIC algorithm is  $NP$  hard. Nevertheless, we wanted to estimate the GW in the finer parcellation of Brainnetome246. We therefore developed an approximation for estimating the GW, which computes the intersection of the top  $G_{in}$  regions across rest and seven tasks. This is only an approximation, since it does not conform fully to our definition of the GW, similar to a 'club' of functional hubs being characterized by a tendency to be more densely functionally connected among themselves than to other brain regions from which they receive integrative information. Nevertheless, this rough approximation shows an overlap with our FRIC algorithm, described in detail above.

Importantly, it should be noted that our definition of FRIC is different from the graph-theoretical concept of rich club, so we compared FRIC with the standard anatomical rich club for the SC<sup>5,7</sup>. Instead of applying the graph-theoretical rich club concept directly to the directed NDTE matrix, we designed the FRIC algorithm. Indeed, in graph theory, assortativity is the tendency of nodes of similar degree to connect to each other, while the concept of rich club measures the density of connectivity between high-degree nodes in an anatomical brain network<sup>14</sup>. The FRIC framework is an even more specific concept, designed to find our proposed quantification of the GW, namely a core set of brain regions, being similar to a 'club' of functional hubs characterized by a tendency to be more densely functionally connected among themselves than to other brain regions from where they receive integrative information and only give out sparsely (see Eq. 11). Still, future research could focus on improving this algorithm even further and investigate the temporal stability of FRIC and, for example, whether multiple clubs could co-exist over time.

**Defining feedforward and feedback (FF) organizational hierarchy using GLM.** In the literature, hierarchy has traditionally been proposed using neuroanatomy and specifically the anatomical hierarchical ordering based on the different structure of feedforward and feedback connections across the brain<sup>1,15</sup>. On this basis, Markov and colleagues defined a framework for assigning hierarchical values to each region in such a way that the difference of the hierarchical values in two brain regions predicts the fraction of feedforward connections coupling those two brain regions<sup>3,41</sup>. In fact, they proposed to use the supragranular layer neurons (SLN) index, defined as the fraction of projections originated in the supragranular layer of the source area to the target area divided by the total number of projections between the SLN of projections. This idea was based on the observations of Felleman and Van Essen<sup>1</sup> and Barone and colleagues<sup>115</sup> that, in the visual system, feedforward projections directed from early visual areas to higher-order areas tend to originate in the supragranular layers of the cortex and terminate in layer 4, whereas projections from higher-order areas to early visual areas originate in the infragranular layers and terminate outside of layer 4.

Our framework for computing the NDTE flow (see above) allows us to extend these seminal ideas to the functional level. Instead of using the anatomically based SLN index, we can use the fraction of functional feedforward connections with respect to the total number of connections, feedforward and feedback (FF), between two brain regions. Consequently, in this FF framework, we assign a functional hierarchical value  $H$  to each brain region such that the difference of the corresponding values between two brain regions predicts the functional fraction of direct connections from one source region  $j$  to another target region  $i$ . We use a GLM to establish this prediction, as follows:

$$\frac{C_{All_j}}{C_{All_j} + C_{All_i}} = g^{-1}(H_i - H_j) \quad (12)$$

where the left-hand side is the fraction of feedforward connections with respect to the total,  $g^{-1}$  is a logistic regression function (which corresponds to fitting a GLM with a binomial family<sup>116</sup>) and  $H_i$  and  $H_j$  are the functional hierarchical values inferred for brain regions  $i$  and  $j$ . To establish a reference point, we assigned a hierarchy value of zero to the last parcel.

As shown in Extended Data Fig. 8, the FF hierarchy measure is similar to  $G_{out}$  in that it is not discriminatory between the seven tasks. On the other hand, it is only weakly correlated with  $G_{in}$ , the integrative measure of incoming information flow needed for establishing FRIC. This is to be expected given that the FF hierarchy

is an ordered measure of the fraction of feedforward and feedback connections, rather than a characterization of a given brain region in terms of incoming and outgoing information flow to/from the other regions in the brain, which is what is needed to establish a meaningful functional hierarchical organization. The results demonstrate that the FF hierarchy measure is not as useful for establishing the full hierarchical organization compared with the proposed NDTE framework.

**Anatomical rich club.** The members of the anatomical rich club were determined using the standard method with graph-theoretical tools<sup>117</sup>. A set of brain regions in the network shows a rich club structure if the level of interconnectivity exceeds the level of connectivity that can be expected on the basis of chance alone. The weighted rich club coefficient  $\Phi(k)$  is computed as the ratio between (1) the weights of connections present within the subnetwork  $S$  of regions with degree  $>k$  and (2) the total sum of weights present within a subset of the same size of the top ranking connections in the network. The normalized rich club coefficient  $\Phi_{norm}(k)$  is computed by dividing  $\Phi(k)$  by  $\Phi_{random}(k)$ , with  $\Phi_{random}(k)$  computed as the average rich club coefficient for each  $k$  of a set of 1,000 randomized graphs (acquired by randomizing the adjacency matrix  $A$  while preserving the degree sequence)<sup>118</sup>. For a given region, membership to the rich club is determined if the normalized rich club coefficient  $\Phi_{norm}(k)$  is larger than one, for a range of increasing degree level  $k$ .

**Whole-brain model of NDTE flow.** The main aim of whole-brain modelling is to infer the dynamical mechanisms generating the observed empirical spatiotemporal dynamics. Here, we would like to estimate the generators underlying the empirical spatiotemporal dynamics in terms of the statistical relationships between the different brain regions, that is, the NDTE flow. Specifically, the whole-brain model will link the structural anatomy (given by the dMRI data) with the functional dynamics (given by the fMRI data) by adapting the free parameters, that is, the generators, to provide the optimal fit between the simulated and empirical NDTE flow. The generators are internal parameters describing the local dynamics of a brain region, such as noise and latency, as well as the strength of the synaptic conductivity of connections between different brain areas linked by the anatomical fibres. We call the matrix of the generative conductivities of the existing anatomical fibres the GABC.

The GABC matrix is estimated from the NDTE flow, which is bidirectional, and is therefore asymmetric. This is unlike the anatomical matrix extracted by dMRI tractography, which is unidirectional. Importantly, GABC is defined as the generators accounting for the causal influence of one neural system over another.

While there is a large literature on dynamic causal modelling of effective connectivity within and among local neuronal (mass) models, the non-linear and emergent dynamical properties of these systems have yet to be explored thoroughly. Previous models used the unidirectional SC to reproduce functional connectivity (FC)<sup>66,119</sup> and a common global conductivity value, meaning that only the scaling factor is optimized. In other words, the scaling factor is one global coupling parameter expressing the conductivity of all fibres equally. In contrast, another possibility is to tune network connections individually (but not independently), which requires a dedicated estimation procedure<sup>120</sup>. This corresponds to the concept of effective connectivity (EC)<sup>21</sup>, which describes how network nodes excite or inhibit each other for a given model of local dynamics. EC thus describes causal interactions whose effects are modulated by the local dynamic regime of the node, which may shape FC in a complex fashion<sup>122</sup>: two areas may be significantly correlated (in FC) although disconnected (zero EC) when strong indirect pathways connect them (that is, large network effect), as was shown by Robinson<sup>123</sup>. Biologically, EC measures the strengths of connections, which depend not only on anatomical properties embodied in SC values (connection densities) but also on heterogeneities in synaptic receptors or neurotransmitters. Gilson and colleagues provided a solution for estimating EC from fMRI FC with information about the directed connectivity at the whole-brain level (divided in a parcellation of 70 brain regions with a couple of thousand connections to tune).

Until now, there has not been a generative model that can explain the bidirectional flow of information across the whole brain, which is non-trivial given that the SC and NDTE matrices are not strongly correlated. Here, we extend existing models by using a more powerful bidirectional measure, namely the NDTE flow, that captures the underlying spatiotemporal dynamical mechanisms, and thus not correlations or time-shifted correlations. For that, we use a recent successful model, namely the Hopf whole-brain model<sup>17</sup>, in combination with a powerful non-gradient-based global optimization algorithm, namely particle swarm optimization.

Briefly in the following, we describe how whole-brain models aim to obtain a balance between complexity and realism to describe the most important features of the brain *in vivo*<sup>124</sup>. This balance is extremely difficult to achieve because of the astronomical number of neurons and the underspecified connectivity at the neural level. The emerging collective macroscopic dynamics of brain models use mesoscopic top-down approximations of brain complexity with dynamical networks of local brain area attractor networks<sup>69,125</sup>. Essentially, these models link anatomical structure (given by the dMRI tractography) and functional dynamics (typically measured with fMRI) to reproduce the whole-brain empirical data<sup>71,126–128</sup>.

Biologically informed models have been useful for modelling neuroimaging data, for example, MEG data<sup>129</sup> and working points near the Hopf bifurcation.

In fact, the emergence of functional brain networks from the combination of local dynamics and structural topology has also been captured by other models, including random walkers<sup>130</sup>, wave equations on a network<sup>131</sup> and epidemic spreading models just beyond the critical epidemic threshold for analysing the information flow<sup>132</sup>.

Here, we use the Hopf whole-brain model consisting of coupled dynamical units (regions of interest or nodes) representing the  $N$  cortical and subcortical brain areas in a given parcellation<sup>75</sup>. The local dynamics of each brain region is described by the normal form of a supercritical Hopf bifurcation, also known as the Landau–Stuart oscillator, which is the canonical model for studying the transition from noisy to oscillatory dynamics<sup>133</sup>. Coupled together with the brain network architecture, the complex interactions between Hopf oscillators have been shown to reproduce important features of brain dynamics observed in electrophysiology<sup>134,135</sup>, MEG<sup>136</sup> and fMRI<sup>137,138</sup>.

The dynamics of an uncoupled brain region  $n$  is given by the following set of coupled dynamical equations, which describe the normal form of a supercritical Hopf bifurcation in Cartesian coordinates:

$$\frac{dx_n}{dt} = [a_n - x_n^2 - y_n^2]x_n - \omega_n y_n + \beta \eta_n(t) \quad (13)$$

$$\frac{dy_n}{dt} = [a_n - x_n^2 - y_n^2]y_n + \omega_n x_n + \beta \eta_n(t) \quad (14)$$

where  $\eta_n(t)$  is additive Gaussian noise with standard deviation  $\beta$ . This normal form has a supercritical bifurcation  $a_n=0$ , so that, if  $a_n>0$ , the system engages in a stable limit cycle with frequency  $f_n=\omega_n/2\pi$ . On the other hand, when  $a_n<0$ , the local dynamics are in a stable fixed point representing a low-activity noisy state. Within this model, the intrinsic frequency  $\omega_n$  of each region is in the 0.01–0.2 Hz band ( $n=1, \dots, N$ ), where  $N$  is the total number of regions. It has been shown that the optimized  $a_n$  form a distribution of negative and positive values, consistent with the heterogeneous mixture of regions at bifurcation, steady state or limit cycle, arising from the optimization of the power spectrum of the regions<sup>75,76</sup>.

We estimated the intrinsic frequencies from the empirical data, as given by the averaged peak frequency of the narrowband BOLD signals of each brain region. The variable  $x_n$  emulates the BOLD signal of each region  $n$ . To model the whole-brain dynamics, we added an additive coupling term representing the input received in region  $n$  from every other region  $p$ , which is weighted by the corresponding SC. The whole-brain dynamics was defined by the following set of coupled equations:

$$\begin{aligned} \frac{dx_n}{dt} = & [a_n - x_n^2 - y_n^2]x_n - \omega_n y_n \\ & + G \sum_{p=1}^N C_{np}(x_p(t-\tau) - x_n) + \beta_n \eta_n(t) \end{aligned} \quad (15)$$

$$\begin{aligned} \frac{dy_n}{dt} = & [a_n - x_n^2 - y_n^2]y_n + \omega_n x_n \\ & + G \sum_{p=1}^N C_{np}(y_p(t-\tau) - y_n) + \beta_n \eta_n(t) \end{aligned} \quad (16)$$

where  $G$  denotes the global coupling weight, scaling equally the total input received in each brain area, and  $\tau$  is a time lag. The SC matrix  $C_{np}$  is estimated and normalized from dMRI tractography (with a maximum of 0.2) and thus symmetric.

During optimization, the strength of connections in  $C_{np}$  is updated on the basis of the fit between the model output and the empirical NDTE flow matrix in terms of correlation. The empirical NDTE matrix is bidirectional and thus asymmetric. Hence, when updating the structural matrix, this will become asymmetric too. Thus,  $C_{np}$  is the GABIC.

We used a global optimization routine of MATLAB, namely the particle swarm optimizer. Particle swarm is a population-based algorithm, similar to the genetic algorithm, which has been widely used in treating ill-structured continuous, constrained as well as unconstrained function optimization problems<sup>139–141</sup>. A population of individuals (called particles) diffuse throughout the searching region of parameters, not dissimilar to swarming insects. At each step, the algorithm evaluates the objective function for each particle. In our case, the objective consisted of maximizing the correlation between the empirical and simulated NDTE matrix (that is, considering the causality between all pairs) by optimizing sequentially the noise level,  $\beta_n$ , the time lag,  $\tau$ , the local bifurcation parameters,  $a_n$ , and most importantly the matrix  $C_{np}$ . The initial values of noise were fixed thus:  $\beta=0.02$ ,  $a_n=-0.02$ , with the time lags also initialized to zero. The diffusion of the particles is optimally adapted by the algorithm to converge to a global maximum.

We remark that the whole-brain model generation of the NDTE matrix achieved here is a non-trivial problem, which is the result of using a non-linear model of a whole-brain coupling structure (SC) that does not correlate with the desired directional information flow (NDTE). This is essential for identifying the underlying GABIC matrix that can generate and mechanistically explain the emergence of the NDTE information flow—as well as other emergent properties such as FRIC.

An important caveat to the findings presented here is the seminal work by Judea Pearl in his book *Causality*<sup>142</sup>, where he shows that any framework of causal

inference is based on inferring causal structures that are equivalent in terms of the probability distributions they generate; that is, they are indistinguishable from observational data, and could only be distinguished by manipulating the whole system.

Nevertheless, our NDTE framework—built from the probabilistic principles of mutual information as shown in the existing literature<sup>15–27,109</sup>—is well suited for our stated aim of determining the functional hierarchical organization of brain function by determining the directional causality.

In line with Pearl's argument, we remark that the NDTE framework does capture the directionality of information flow needed for our purposes, but not the generative underlying physical coupling. For this we built the whole-brain GABIC model, which is able to capture the brain physical generators accounting for the causal influence of one neural system over another. We note that another promising framework for capturing the generators could be the dynamical causal modelling (DCM) framework<sup>143</sup>. However, it should be noted that DCM is currently limited to applications with very limited number of areas (at most approximately 10), although recent work has tried to apply the framework to the whole brain<sup>144,145</sup>.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The multimodal neuroimaging data are freely available from HCP.

## Code availability

The code used to run the analysis from the HCP data is available on GitHub (<https://github.com/decolab/nhb-ndte>).

Received: 18 March 2020; Accepted: 12 October 2020;

Published online: 4 January 2021

## References

1. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
2. Mesulam, M. M. From sensation to cognition. *Brain J. Neurol.* **121**, 1013–1052 (1998).
3. Markov, N. T. et al. Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.* **522**, 225–259 (2014).
4. Bullmore, E. & Sporns, O. The economy of brain network organization. *Nat. Rev. Neurosci.* **13**, 336–349 (2012).
5. van den Heuvel, M. P. & Sporns, O. Rich-club organization of the human connectome. *J. Neurosci.* **31**, 15775–15786 (2011).
6. Hagmann, P. et al. Mapping the structural core of human cerebral cortex. *PLoS Biol.* **6**, e159 (2008).
7. Zamora-Lopez, G., Zhou, C. & Kurths, J. Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. *Front. Neuroinform.* **4**, 1 (2010).
8. Margulies, D. S. et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl Acad. Sci. USA* **113**, 12574–12579 (2016).
9. Huntenburg, J. M., Bazin, P. L. & Margulies, D. S. Large-scale gradients in human cortical organization. *Trends Cogn. Sci.* **22**, 21–31 (2018).
10. Buckner, R. L. & DiNicola, L. M. The brain's default network: updated anatomy, physiology and evolving insights. *Nat. Rev. Neurosci.* **20**, 593–608 (2019).
11. Atasoy, S., Donnelly, I. & Pearson, J. Human brain networks function in connectome-specific harmonic waves. *Nat. Commun.* **7**, 10340 (2016).
12. Atasoy, S. et al. Connectome-harmonic decomposition of human brain activity reveals dynamical repertoire re-organization under LSD. *Sci. Rep.* **7**, 17661 (2017).
13. Norman, D. A. & Shallice, T. in *Consciousness and Self-Regulation* (eds Davidson, R. J., Schwartz, G. E. & Shapiro, D.) 1–18 (Plenum Press, 1980).
14. Baars, B. J. *A Cognitive Theory of Consciousness* (Cambridge Univ. Press, 1989).
15. Dehaene, S., Kerszberg, M. & Changeux, J. P. A neuronal model of a global workspace in effortful cognitive tasks. *Proc. Natl Acad. Sci. USA* **95**, 14529–14534 (1998).
16. Honey, C. J. & Sporns, O. Dynamical consequences of lesions in cortical networks. *Hum. Brain Mapp.* **29**, 802–809 (2008).
17. Alstott, J., Breakspear, M., Hagmann, P., Cammoun, L. & Sporns, O. Modeling the impact of lesions in the human brain. *PLoS Comput. Biol.* **5**, e1000408 (2009).
18. Glasser, M. F. et al. The human connectome project's neuroimaging approach. *Nat. Neurosci.* **19**, 1175–1187 (2016).
19. Van Essen, D. C. et al. The WU-minn human connectome project: an overview. *NeuroImage* **80**, 62–79 (2013).

20. Lobier, M., Siebenhuhner, F., Palva, S. & Palva, J. M. Phase transfer entropy: a novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *NeuroImage* **85**, 853–872 (2014).
21. Hillebrand, A. et al. Direction of information flow in large-scale resting-state networks is frequency-dependent. *Proc. Natl Acad. Sci USA* **113**, 3867–3872 (2016).
22. Barnett, L., Muthukumaraswamy, S. D., Carhart-Harris, R. L. & Seth, A. K. Decreased directed functional connectivity in the psychedelic state. *NeuroImage* **209**, 116462 (2020).
23. Quiroga, R. Q., Kraskov, A., Kreuz, T. & Grassberger, P. Performance of different synchronization measures in real data: a case study on electroencephalographic signals. *Phys. Rev. E* **65**, 041903 (2002).
24. Stouffer, S. A., Suchman, E. A., Devinney, L. C., Star, S. A. & Williams, R. M. Jr *The American Soldier: Adjustment during Army Life (Studies in Social Psychology in World War II)* 1 (Princeton Univ. Press, 1949).
25. Brovelli, A., Chicharro, D., Badler, J. M., Wang, H. & Jirsa, V. Characterization of cortical networks and corticocortical functional connectivity mediating arbitrary visuomotor mapping. *J. Neurosci.* **35**, 12643–12658 (2015).
26. Quiroga, R. Q., Arnhold, J. & Grassberger, P. Learning driver-response relationships from synchronization patterns. *Phys. Rev. E* **61**, 5142–5148 (2000).
27. Quiroga, R. Q., Kreuz, T. & Grassberger, P. Event synchronization: a simple and fast method to measure synchronicity and time delay patterns. *Phys. Rev. E* **66**, 041904 (2002).
28. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
29. Fan, L. et al. The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).
30. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).
31. Klein, A. & Tourville, J. 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* **6**, 171 (2012).
32. Glasser, M. F. & Van Essen, D. C. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* **31**, 11597–11616 (2011).
33. Beer, A. et al. Tissue damage within normal appearing white matter in early multiple sclerosis: assessment by the ratio of T1- and T2-weighted MR image intensity. *J. Neurol.* **263**, 1495–1502 (2016).
34. Arshad, M., Stanley, J. A. & Raz, N. Test-retest reliability and concurrent validity of in vivo myelin content indices: myelin water fraction and calibrated T1w/T2w image ratio. *Hum. Brain Mapp.* **38**, 1780–1790 (2017).
35. Righart, R. et al. Cortical pathology in multiple sclerosis detected by the T1/T2-weighted ratio from routine magnetic resonance imaging. *Ann. Neurol.* **82**, 519–529 (2017).
36. Cooper, G. et al. Standardization of T1w/T2w ratio improves detection of tissue damage in multiple sclerosis. *Front. Neurol.* **10**, 334 (2019).
37. Burt, J. B. et al. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **21**, 1251–1259 (2018).
38. Demirtas, M. et al. Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. *Neuron* **101**, 1181–1194 e1113 (2019).
39. Hunt, B. A. et al. Relationships between cortical myeloarchitecture and electrophysiological networks. *Proc. Natl Acad. Sci USA* **113**, 13510–13515 (2016).
40. Barch, D. M. et al. Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage* **80**, 169–189 (2013).
41. Markov, N. T. et al. Cortical high-density counterstream architectures. *Science* **342**, 1238406 (2013).
42. Poldrack, R. A. & Gorgolewski, K. J. Making big data open: data sharing in neuroimaging. *Nat. Neurosci.* **17**, 1510–1517 (2014).
43. Alfaro-Almagro, F. et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400–424 (2018).
44. Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).
45. Squire, L. R., Stark, C. E. & Clark, R. E. The medial temporal lobe. *Annu. Rev. Neurosci.* **27**, 279–306 (2004).
46. Eichenbaum, H., Yonelinas, A. P. & Ranganath, C. The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci.* **30**, 123–152 (2007).
47. Berridge, K. C. & Robinson, T. E. Parsing reward. *Trends Neurosci.* **26**, 507–513 (2003).
48. Berridge, K. C. & Kringelbach, M. L. Affective neuroscience of pleasure: reward in humans and animals. *Psychopharmacology* **199**, 457–480 (2008).
49. Haber, S. N. & Knutson, B. The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* **35**, 4–26 (2010).
50. Di Martino, A. et al. Functional connectivity of human striatum: a resting state fMRI study. *Cereb. Cortex* **18**, 2735–2747 (2008).
51. Zald, D. H. & Pardo, J. V. Emotion, olfaction, and the human amygdala: amygdala activation during aversive olfactory stimulation. *Proc. Natl Acad. Sci USA* **94**, 4119–4124 (1997).
52. Swanson, L. W. & Petrovich, G. D. What is the amygdala? *Trends Neurosci.* **21**, 323–331 (1998).
53. Schoenbaum, G., Setlow, B., Saddoris, M. P. & Gallagher, M. Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* **39**, 855–867 (2003).
54. LeDoux, J. The emotional brain, fear, and the amygdala. *Cell. Mol. Neurobiol.* **23**, 727–738 (2003).
55. Margulies, D. S. et al. Precuneus shares intrinsic functional architecture in humans and monkeys. *Proc. Natl Acad. Sci USA* **106**, 20069–20074 (2009).
56. Cavanna, A. E. & Trimble, M. R. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain J. Neurol.* **129**, 564–583 (2006).
57. Northoff, G. & Bermpohl, F. Cortical midline structures and the self. *Trends Cogn. Sci.* **8**, 102–107 (2004).
58. Mesulam, M. M. Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philos. Trans. R. Soc. Lond. Ser. B* **354**, 1325–1346 (1999).
59. Paus, T. Primate anterior cingulate cortex: where motor control, drive and cognition interface. *Nat. Rev. Neurosci.* **2**, 417–424 (2001).
60. Leech, R. & Sharp, D. J. The role of the posterior cingulate cortex in cognition and disease. *Brain J. Neurol.* **137**, 12–32 (2014).
61. Laureys, S., Owen, A. M. & Schiff, N. D. Brain function in coma, vegetative state, and related disorders. *Lancet Neurology* **3**, 537–546 (2004).
62. Dehaene, S. & Changeux, J. P. Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS Biol.* **3**, e141 (2005).
63. Deco, G., Cruzat, J. & Kringelbach, M. L. Brain songs framework for discovering the relevant timescale of the human brain. *Nat. Commun.* **10**, 583 (2019).
64. Glomb, K. et al. Functional harmonics reveal multi-dimensional basis functions underlying cortical organization. Preprint at *bioRxiv* <https://doi.org/10.1101/699678> (2019).
65. Eickhoff, S. B., Constable, R. T. & Yeo, B. T. T. Topographic organization of the cerebral cortex and brain cartography. *NeuroImage* **170**, 332–347 (2018).
66. Deco, G. & Kringelbach, M. L. Great expectations: using whole-brain computational connectomics for understanding neuropsychiatric disorders. *Neuron* **84**, 892–905 (2014).
67. Berridge, K. C. & Kringelbach, M. L. Pleasure systems in the brain. *Neuron* **86**, 646–664 (2015).
68. Kringelbach, M. L. The orbitofrontal cortex: linking reward to hedonic experience. *Nat. Rev. Neurosci.* **6**, 691–702 (2005).
69. Breakspear, M. Dynamic models of large-scale brain activity. *Nat. Neurosci.* **20**, 340–352 (2017).
70. Kringelbach, M. L. & Deco, G. Brain states and transitions: insights from computational neuroscience. *Cell Rep.* **32**, 108128 (2020).
71. Deco, G. et al. Whole-brain multimodal neuroimaging model using serotonin receptor maps explains non-linear functional effects of LSD. *Curr. Biol.* **28**, 3065–3074 (2018).
72. Honey, C. J., Kötter, R., Breakspear, M. & Sporns, O. Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl Acad. Sci USA* **104**, 10240–10245 (2007).
73. Deco, G. et al. Awakening: predicting external stimulation forcing transitions between different brain states. *Proc. Natl Acad. Sci USA* **116**, 18088–18097 (2019).
74. Deco, G., Van Harteveld, T., Fernandes, H. M., Stevner, A. B. A. & Kringelbach, M. L. The most relevant human brain regions for functional connectivity: evidence for a dynamical workspace of binding nodes from whole-brain computational modelling. *NeuroImage* **146**, 197–210 (2017).
75. Deco, G., Kringelbach, M. L., Jirsa, V. & Ritter, P. The dynamics of resting fluctuations in the brain: metastability and its dynamical core. *Sci. Rep.* **7**, 3095 (2017).
76. Saenger, V. M. et al. Uncovering the underlying mechanisms and whole-brain dynamics of therapeutic deep brain stimulation for Parkinson's disease. *Sci. Rep.* **7**, 9882 (2017).
77. Donnelly-Kehoe, P. D. et al. Reliable local dynamics in the brain across sessions are revealed by whole-brain modelling of resting state activity. *Hum. Brain Mapp.* **40**, 2967–2980 (2019).
78. Deco, G. et al. Single or multi-frequency generators in on-going MEG data: a mechanistic whole-brain model of empirical MEG data. *NeuroImage* **152**, 538–550 (2017).
79. Kuramoto, Y. *Chemical Oscillations, Waves, and Turbulence* (Springer-Verlag, 1984).
80. Tononi, G., Sporns, O. & Edelman, G. M. A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl Acad. Sci USA* **91**, 5033–5037 (1994).

81. Northoff, G. What the brain's intrinsic activity can tell us about consciousness? A tri-dimensional view. *Neurosci. Biobehav. Rev.* **37**, 726–738 (2013).
82. Deco, G. et al. Perturbation of whole-brain dynamics in silico reveals mechanistic differences between brain states. *NeuroImage* **169**, 46–56 (2017).
83. Setsompop, K. et al. Pushing the limits of in vivo diffusion MRI for the Human Connectome Project. *NeuroImage* **80**, 220–233 (2013).
84. Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
85. Li, N. et al. Toward a unified connectomic target for deep brain stimulation in obsessive-compulsive disorder. *Nat. Commun.* **11**, 3364 (2019).
86. Horn, A., Neumann, W. J., Degen, K., Schneider, G. H. & Kuhn, A. A. Toward an electrophysiological 'sweet spot' for deep brain stimulation in the subthalamic nucleus. *Hum. Brain Mapp.* **38**, 3377–3390 (2017).
87. Horn, A. & Blankenburg, F. Toward a standardized structural-functional group connectome in MNI space. *NeuroImage* **124**, 310–322 (2016).
88. Maier-Hein, K. H. et al. The challenge of mapping the human connectome based on diffusion tractography. *Nat. Commun.* **8**, 1349 (2017).
89. Schilling, K. G. et al. Challenges in diffusion MRI tractography – lessons learned from international benchmark competitions. *Magn. Reson. Imaging* **57**, 194–209 (2019).
90. Glasser, M. F. et al. The minimal preprocessing pipelines for the human connectome project. *NeuroImage* **80**, 105–124 (2013).
91. Smith, S. M. et al. Resting-state fMRI in the human connectome project. *NeuroImage* **80**, 144–168 (2013).
92. Navarro Schroder, T., Haak, K. V., Zaragoza Jimenez, N. I., Beckmann, C. F. & Doeller, C. F. Functional topography of the human entorhinal cortex. *eLife* **4**, e06738 (2015).
93. Salimi-Khorshidi, G. et al. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage* **90**, 449–468 (2014).
94. Griffanti, L. et al. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage* **95**, 232–247 (2014).
95. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 156869 (2011).
96. Vidaurre, D. et al. Spontaneous cortical activity transiently organises into frequency specific phase-coupling networks. *Nat. Commun.* **9**, 2987 (2018).
97. Hillebrand, A. & Barnes, G. R. Beamformer analysis of MEG data. *Int. Rev. Neurobiol.* **68**, 149–171 (2005).
98. Kantz, H. & Schreiber, T. *Nonlinear Time Series Analysis* (Cambridge Univ. Press, 1997).
99. Granger, C. Testing for causality. *J. Econ. Dyn. Control* **2**, 329–352 (1980).
100. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464 (2000).
101. Vicente, R., Wibral, M., Lindner, M. & Pipa, G. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **30**, 45–67 (2011).
102. Chicharro, D. & Ledberg, A. Framework to study dynamic dependencies in networks of interacting processes. *Phys. Rev. E* **86**, 041901 (2012).
103. Wibral, M., Vicente, R. & Lindner, M. in *Directed Information Measures in Neuroscience* (eds Wibral, M., Vicente, R. & Lizier, J.) 3–36 (Springer, 2014).
104. Weber, I., Florin, E., von Papen, M. & Timmermann, L. The influence of filtering and downsampling on the estimation of transfer entropy. *PLoS ONE* **12**, e0188210 (2017).
105. Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Doyne Farmer, J. Testing for nonlinearity in time series: the method of surrogate data. *Phys. D* **58**, 77–94 (1992).
106. Diks, C. & Fang, H. Transfer entropy for nonparametric granger causality detection: an evaluation of different resampling methods. *Entropy* **19**, 372 (2017).
107. Hinich, M. J., Mendes, E. M. & Stone, L. Detecting nonlinearity in time series: Surrogate and bootstrap approaches. *Stud. Nonlinear Dyn. Econ.* <https://doi.org/10.2202/1558-3708.1268> (2005).
108. Faes, L., Porta, A. & Nollo, G. Mutual nonlinear prediction as a tool to evaluate coupling strength and directionality in bivariate time series: comparison among different strategies based on  $k$  nearest neighbors. *Phys. Rev. E* **78**, 026201 (2008).
109. Pereda, E., Quiroga, R. Q. & Bhattacharya, J. Nonlinear multivariate analysis of neurophysiological signals. *Prog. Neurobiol.* **77**, 1–37 (2005).
110. Schreiber, T. & Schmitz, A. Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.* **77**, 635–638 (1996).
111. Fisher, R. *Statistical Methods for Research Workers*, 13th ed (Oliver and Boyd, 1925).
112. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
113. Seth, A. K., Chorley, P. & Barnett, L. C. Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. *NeuroImage* **65**, 540–555 (2013).
114. Fornito, A., Zalesky, A. & Bullmore, E. *Fundamentals of Brain Network Analysis* (Academic Press, 2016).
115. Barone, P., Batardi, A., Knoblauch, K. & Kennedy, H. Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. *J. Neurosci.* **20**, 3263–3281 (2000).
116. McCullagh, P. & Nelder, J. A. *Generalized Linear Models*, 2nd edition (Chapman and Hall, 1989).
117. van den Heuvel, M. P. & Sporns, O. An anatomical substrate for integration among functional networks in human cortex. *J. Neurosci.* **33**, 14489–14500 (2013).
118. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* **52**, 1059–1069 (2010).
119. Deco, G., Jirsa, V. K. & McIntosh, A. R. Emerging concepts for the dynamical organization of resting state activity in the brain. *Nat. Rev. Neurosci.* **12**, 43–56 (2011).
120. Gilson, M., Moreno-Bote, R., Ponce-Alvarez, A., Ritter, P. & Deco, G. Estimation of directed effective connectivity from fMRI functional connectivity hints at asymmetries of cortical connectome. *PLoS Comput. Biol.* **12**, e1004762 (2016).
121. Friston, K. J. Functional and effective connectivity: a review. *Brain Connectivity* **1**, 13–36 (2011).
122. Park, H. J. & Friston, K. Structural and functional brain networks: from connections to cognition. *Science* **342**, 1238411 (2013).
123. Robinson, P. A. Interrelating anatomical, effective, and functional brain connectivity using propagators and neural field theory. *Phys. Rev. E* **85**, 011912 (2012).
124. Cabral, J., Kringselbach, M. L. & Deco, G. Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: models and mechanisms. *NeuroImage* **160**, 84–96 (2017).
125. Deco, G. & Jirsa, V. K. Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. *J. Neurosci.* **32**, 3366–3375 (2012).
126. Jirsa, V. K., Jantzen, K. J., Fuchs, A. & Kelso, J. A. S. Spatiotemporal forward solution of the EEG and MEG using network modeling. *IEEE Trans. Med. Imaging* **21**, 493–504 (2002).
127. Deco, G., Tononi, G., Boly, M. & Kringselbach, M. L. Rethinking segregation and integration: contributions of whole-brain modelling. *Nat. Rev. Neurosci.* **16**, 430–439 (2015).
128. Kringselbach, M. L. et al. Dynamic coupling of whole-brain neuronal and neurotransmitter systems. *Proc. Natl Acad. Sci. USA* **117**, 9566–9576 (2020).
129. Tewarie, P. et al. Relationships between neuronal oscillatory amplitude and dynamic functional connectivity. *Cereb. Cortex* **29**, 2668–2681 (2019).
130. Abdelnour, F., Voss, H. U. & Raj, A. Network diffusion accurately models the relationship between structural and functional brain connectivity networks. *NeuroImage* **90**, 335–347 (2014).
131. Caputo, J. G., Khames, I., Knipper, A. & Panayotaros, P. Periodic orbits in nonlinear wave equations on networks. *J. Phys. A* **50**, 375101 (2017).
132. Meier, J. et al. The epidemic spreading model and the direction of information flow in brain networks. *NeuroImage* **152**, 639–646 (2017).
133. Kuznetsov, Y. A. *Elements of Applied Bifurcation Theory* (Springer, 1998).
134. Freyer, F. et al. Biophysical mechanisms of multistability in resting-state cortical rhythms. *J. Neurosci.* **31**, 6353–6361 (2011).
135. Freyer, F., Roberts, J. A., Ritter, P. & Breakspear, M. A canonical model of multistability and scale-invariance in biological systems. *PLoS Comput. Biol.* **8**, e1002634 (2012).
136. Deco, G. et al. Single or multiple frequency generators in on-going brain activity: a mechanistic whole-brain model of empirical MEG data. *NeuroImage* **152**, 538–550 (2017).
137. Kringselbach, M. L., McIntosh, A. R., Ritter, P., Jirsa, V. K. & Deco, G. The rediscovery of slowness: exploring the timing of cognition. *Trends Cogn. Sci.* **19**, 616–628 (2015).
138. Deco, G., Kringselbach, M. L., Jirsa, V. K. & Ritter, P. The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. *Sci. Rep.* **7**, 3095 (2017).
139. Del Valle, Y., Venayagamoorthy, G. K., Mohagheghi, S., Hernandez, J.-C. & Harley, R. G. Particle swarm optimization: basic concepts, variants and applications in power systems. *IEEE Trans. Evol. Comput.* **12**, 171–195 (2008).
140. Kennedy, J. & Eberhart, R. Particle swarm optimization. *Proc. IEEE Int. Conf. Neural Netw.* **4**, 1942–1948 (1995).
141. Sengupta, S., Basak, S. & Peters, R. Particle swarm optimization: a survey of historical and recent developments with hybridization perspectives. *Mach. Learn. Knowl. Extraction* **1**, 157–191 (2019).
142. Pearl, J. *Causality: Models, Reasoning and Inference* (Cambridge Univ. Press, 2009).

143. Friston, K. J., Harrison, L. & Penny, W. Dynamic causal modelling. *NeuroImage* **19**, 1273–1302 (2003).
144. Frassle, S. et al. Regression DCM for fMRI. *NeuroImage* **155**, 406–421 (2017).
145. Prando, G. et al. Sparse DCM for whole-brain effective connectivity from resting-state fMRI data. *NeuroImage* **208**, 116367 (2020).

## Acknowledgements

The authors thank D. Chicharro, R. Quian Quiroga, A. Brovelli, J.-P. Changeux and S. Dehaene for their insightful comments on the manuscript, as well as A. Horn and N. Li for their help with diffusion MRI. G.D. is supported by a Spanish research project (ref. PID2019-105772GB-I00 AEI FEDER EU) funded by the Spanish Ministry of Science, Innovation and Universities (MCIU), State Research Agency (AEI) and European Regional Development Funds (FEDER); HBP SGA3 Human Brain Project Specific Grant Agreement 3 (grant agreement no. 945539), funded by the EU H2020 FET Flagship programme and SGR Research Support Group support (ref. 2017 SGR 1545), funded by the Catalan Agency for Management of University and Research Grants (AGAUR). M.L.K. is supported by the ERC Consolidator Grant: CAREGIVING (no. 615539), Center for Music in the Brain, funded by the Danish National Research Foundation (DNRF117), and Centre for Eudaimonia and Human Flourishing funded by the Pettit and Carlsberg Foundations. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## Author contributions

G.D. and M.L.K. designed the study, developed the method and wrote the manuscript. D.V. contributed methods for the MEG analysis. All the authors edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41562-020-01003-6>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41562-020-01003-6>.

Correspondence and requests for materials should be addressed to G.D. or M.L.K.

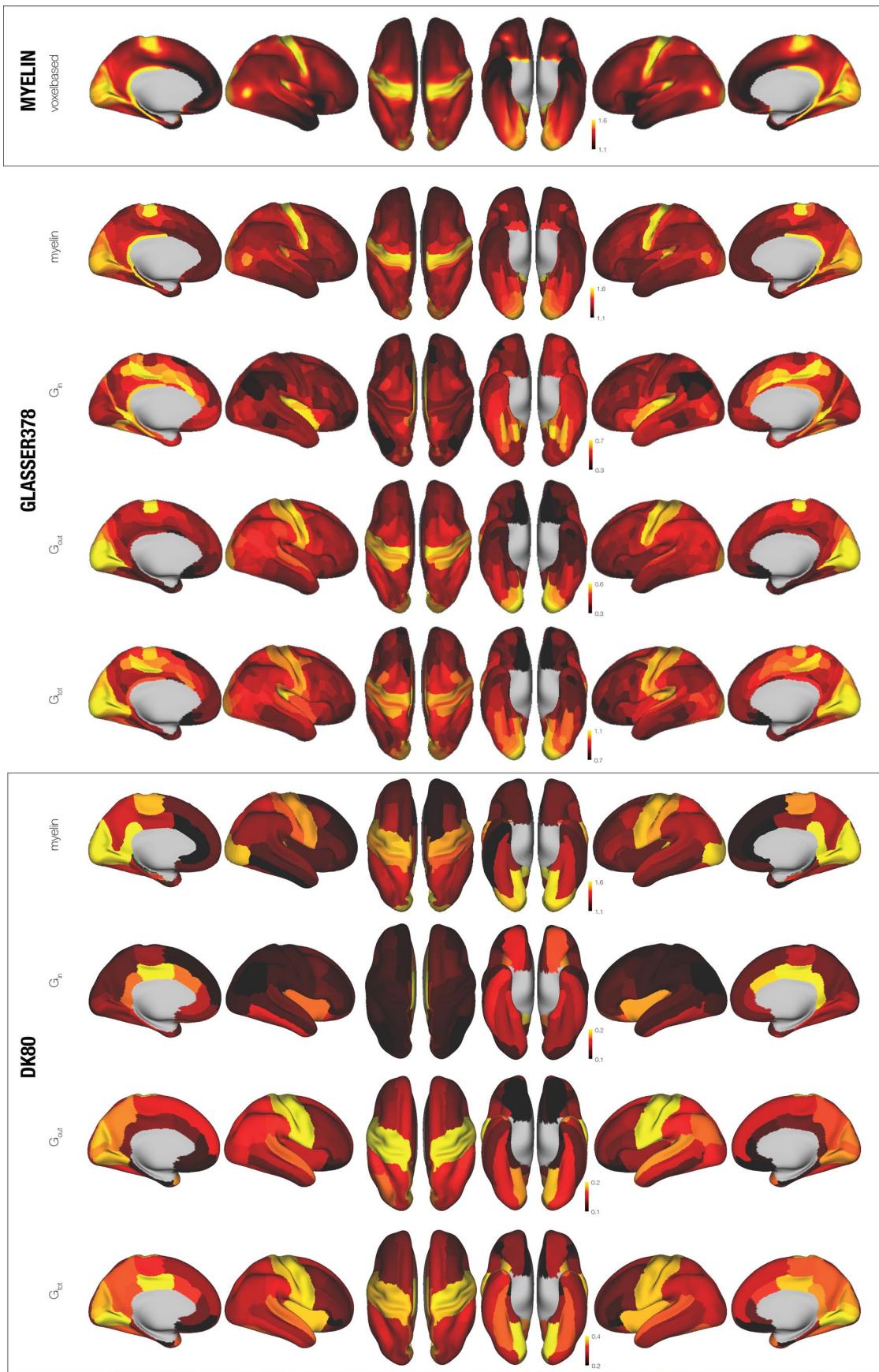
Peer review information *Nature Human Behaviour* thanks Tianzi Jiang, Preejas Tewarie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Marike Schiffer.

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

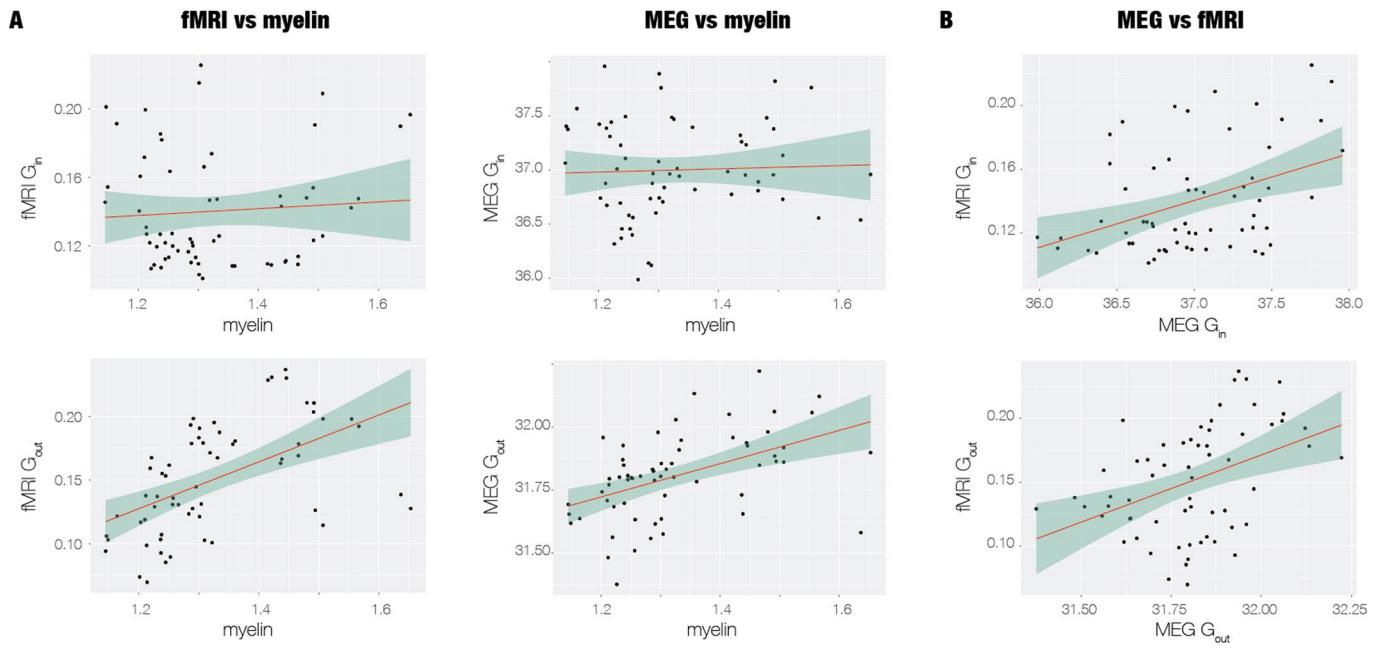


**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>. © The Author(s) 2021

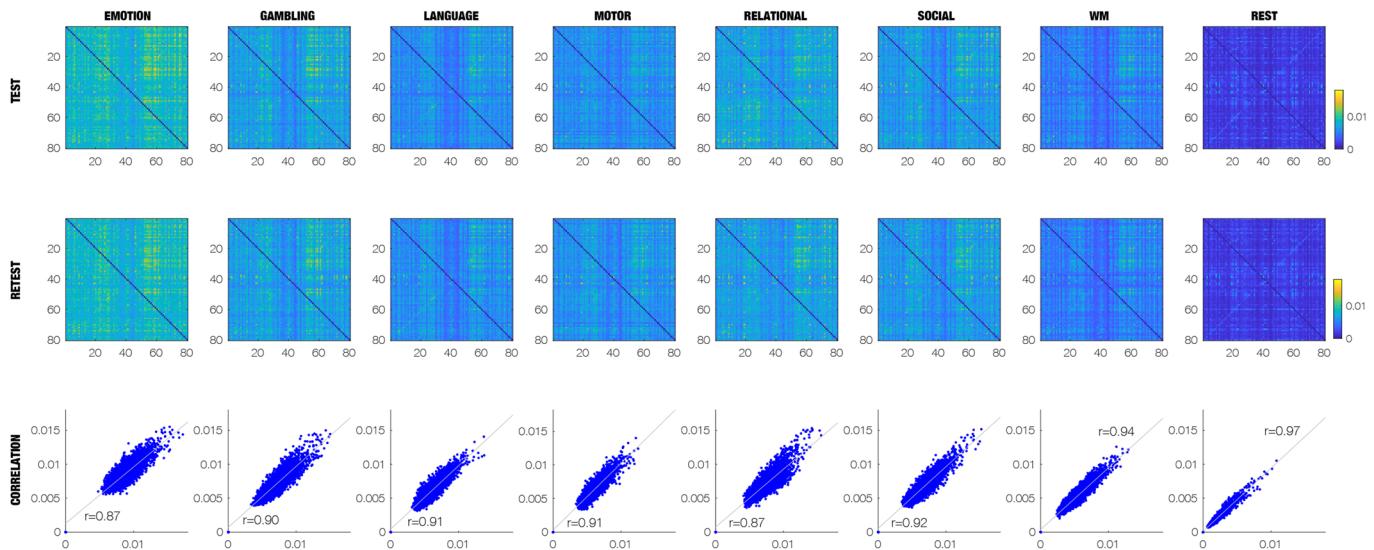


Extended Data Fig. 1 | See next page for caption.

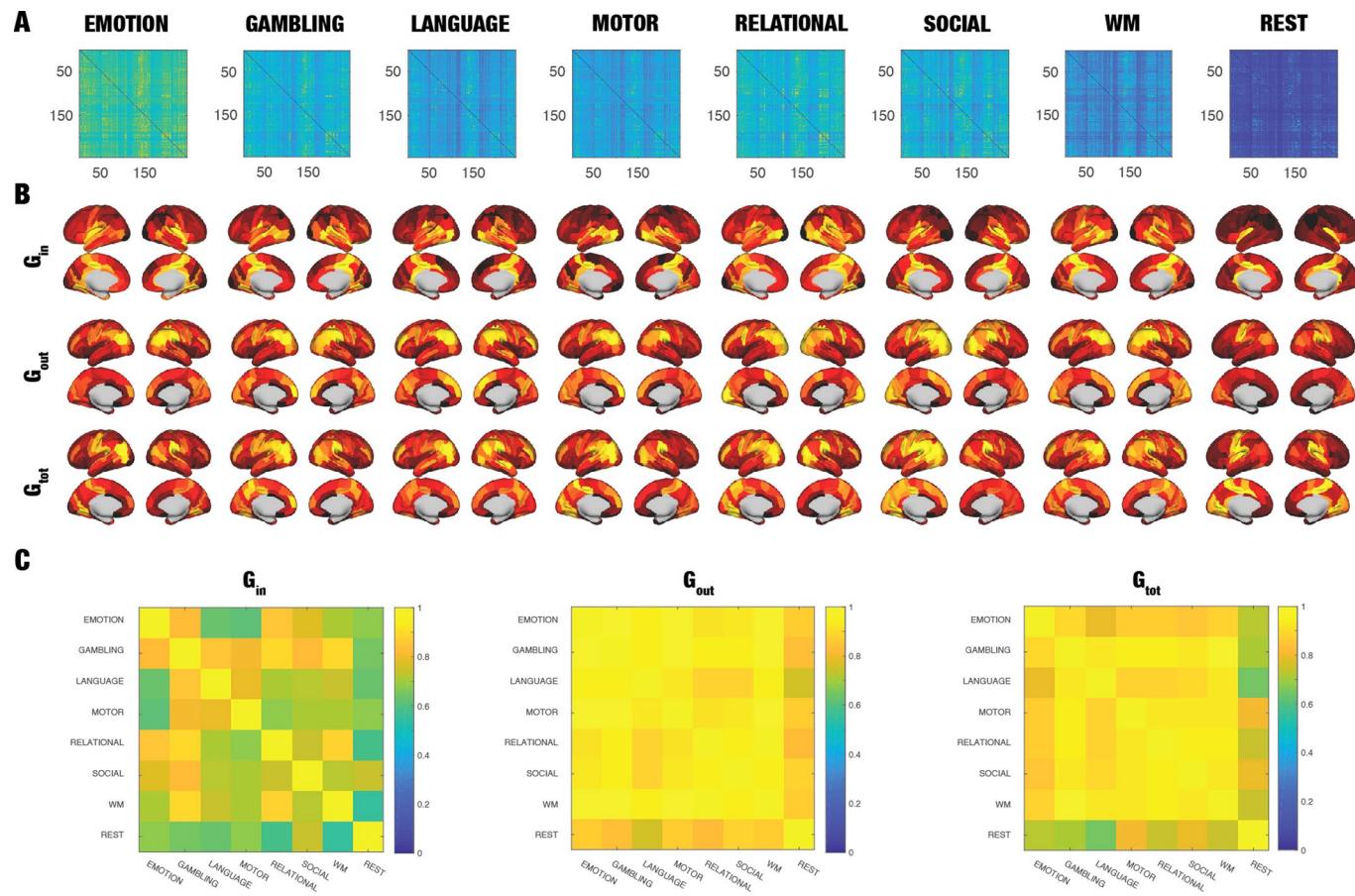
**Extended Data Fig. 1 | Functional hierarchy of resting state found using NDTE rendered on Glasser and DK80 parcellations.** At the top is shown the full cortical voxelbased renderings of myelin (T1w/T2w). Below is shown renderings for both the Glasser and DK80 parcellations of myelin and the NDTE measures ( $G_{inr}$ ,  $G_{out}$  and  $G_{tot}$ ) in 1003 HCP participants.



**Extended Data Fig. 2 | Validating NDTE framework using MEG from 89 HCP participants.** We used the NDTE framework on MEG resting state data from 89 HCP participants (each having three resting state sessions) and extracted the timeseries from the 62 cortical regions of the DK80 parcellation (see Methods). On the left is shown the scatterplots for  $G_{in}$  and  $G_{out}$  for fMRI resting state data versus myelin (T1w/T2w). On the right is shown the same scatterplots for  $G_{in}$  and  $G_{out}$  for MEG resting state data versus myelin. The Pearson correlation values are  $r(124)=0.04$  for  $G_{in}$  (14–22 Hz, window size 1000 ms, not significant, 95% CI = -0.21–0.29),  $r(124)=0.48$  for  $G_{out}$  (22.5–30.5 Hz, window size 500 ms,  $P < 0.001$ , 95% CI = 0.26–0.65). B) The figure shows the scatterplots between  $G_{in}$  and  $G_{out}$  for MEG and fMRI data with correlation values of  $r(124)=0.40$  for  $G_{in}$  ( $P < 0.001$ , 95% CI = 0.17–0.59) and  $r(124)=0.41$  for  $G_{out}$  ( $P < 0.001$ , 95% CI = 0.18–0.60).

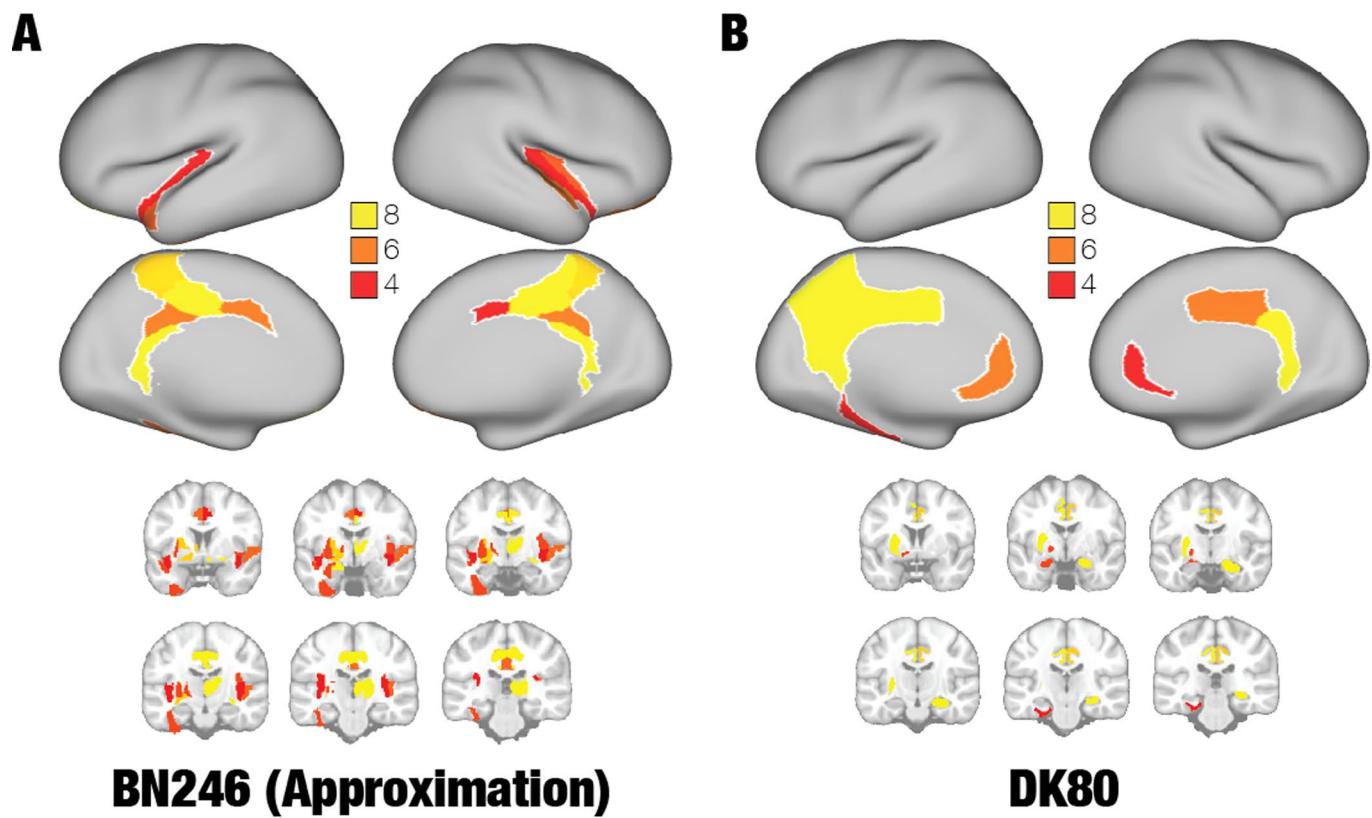


**Extended Data Fig. 3 | Validation of NDTE framework on test-retest data for task and rest.** We validated the NDTE framework on test-retest data and found very high reproducibility (of up to a 0.97 correlation) using only a subsample of 45 participants of the 1003 participants. The correlation values for the rest-retest are the following: EMOTION:  $r(3160)=0.866$ ,  $P < 0.001$ , 95% CI = 0.860–0.872); GAMBLING:  $r(3160)=0.898$ ,  $P < 0.001$ , 95% CI = 0.893–0.903; LANGUAGE:  $r(3160)=0.907$ ,  $P < 0.001$ , 95% CI = 0.902–0.911; MOTOR:  $r(3160)=0.907$ ,  $P < 0.001$ , 95% CI = 0.903–0.912; RELATIONAL:  $r(3160)=0.866$ ,  $P < 0.001$ , 95% CI = 0.860–0.872); SOCIAL:  $r(3160)=0.924$ ,  $P < 0.001$ , 95% CI = 0.920–0.927); WM  $r(3160)=0.939$ ,  $P < 0.001$ , 95% CI = 0.936–0.942); REST:  $r(3160)=0.970$ ,  $P < 0.001$ , 95% CI = 0.968–0.971).

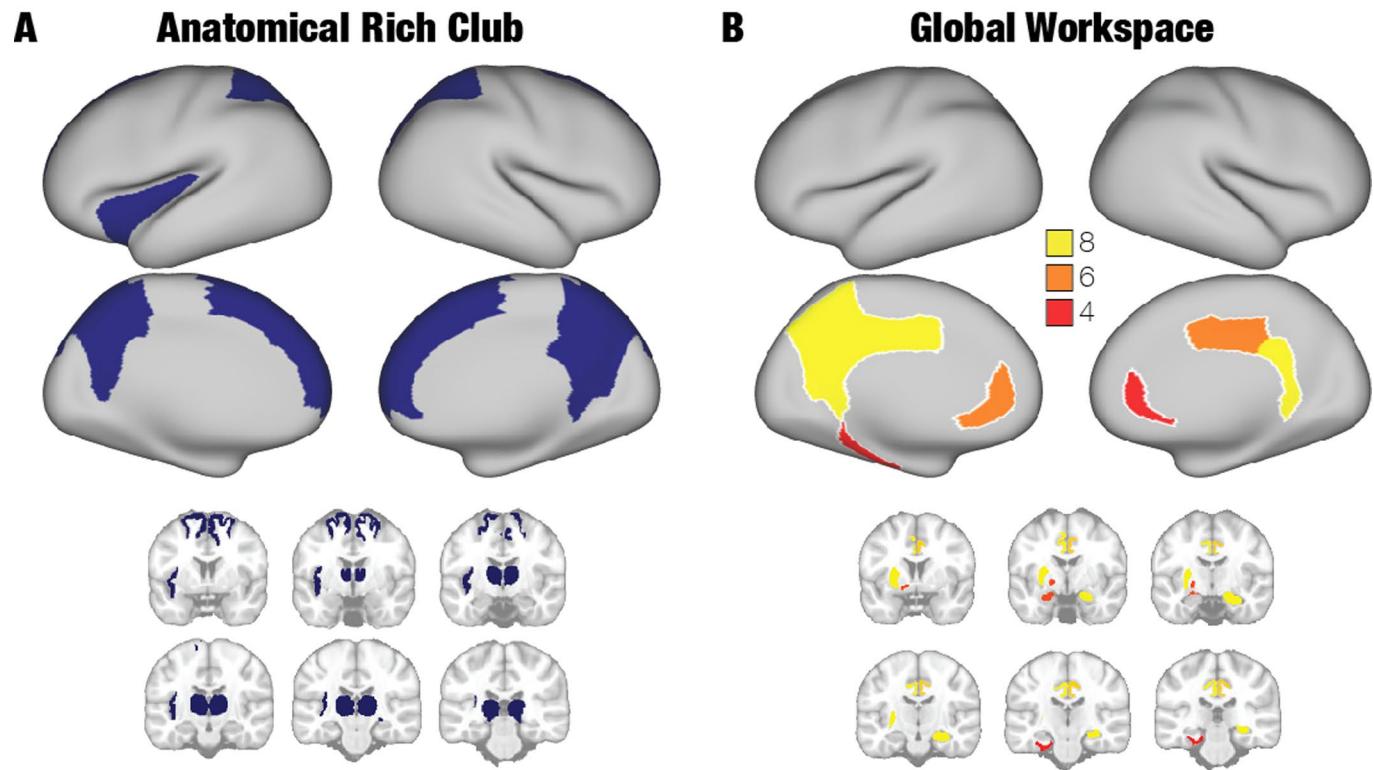


**Extended Data Fig. 4 | Functional hierarchy across different tasks and rest in Brainnetome246 parcellation.** **a**, NDTE matrices for all seven tasks and resting state from all HCP participants in the BN246 parcellation. **b**, Cortical renderings of the incoming ( $G_{in}$ ), outgoing ( $G_{out}$ ) and total ( $G_{tot}$ ) directional information flow computed from the NDTE matrix using 3D views from the side and midline (similar to Fig. 3). **c**, Matrices of the comparison of incoming ( $G_{in}$ ), outgoing ( $G_{out}$ ) and total ( $G_{tot}$ ) directional information flow in the seven tasks and rest. Similar to the results for the DK80 and Glasser378 parcellations, the renderings of the incoming flow of information (receivers) are significantly different between tasks and rest.

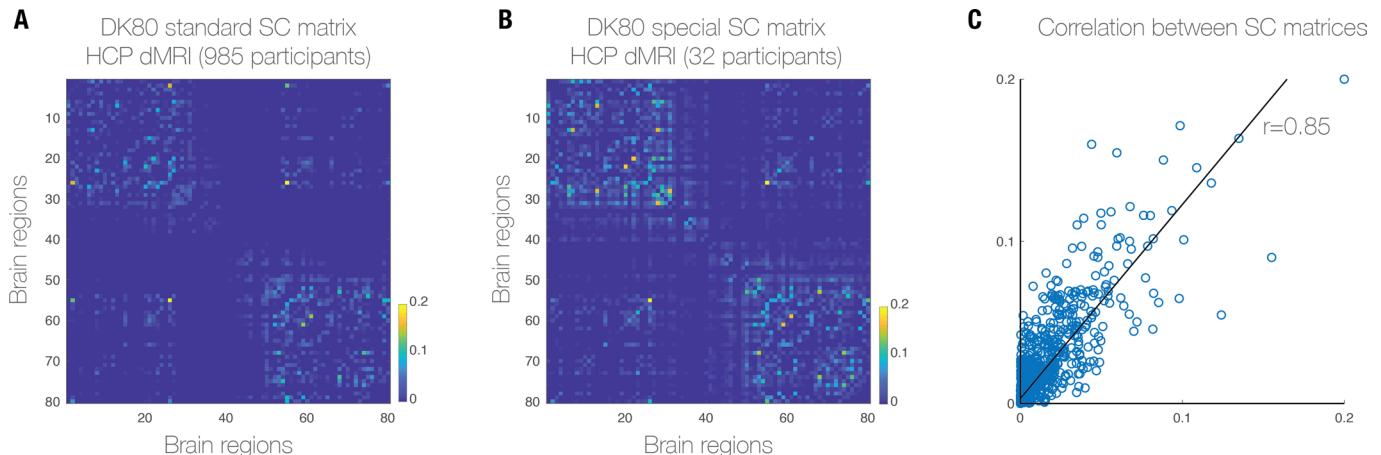
## GLOBAL WORKSPACE (intersection of FRICs)



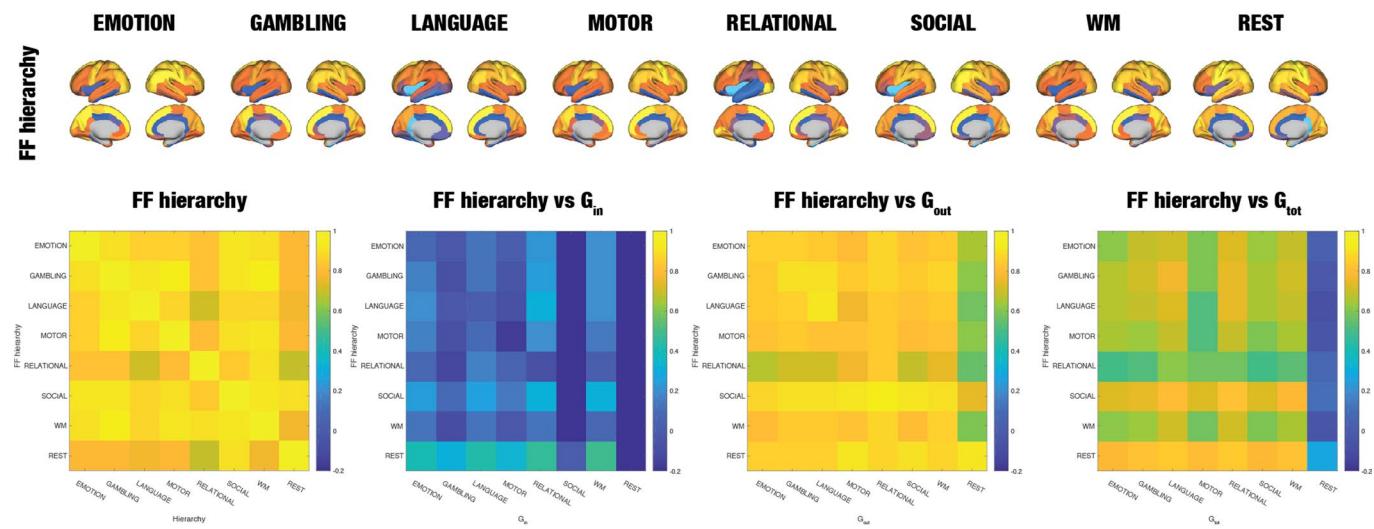
**Extended Data Fig. 5 | Approximation of the Global Workspace in the fine-scale Brainnetome246 parcellation for comparison with formal definition of the Global Workspace in the DK80 parcellation.** **a**, The computational complexity of the FRIC algorithm makes it inappropriate for finer-scale parcellations so we developed an approximation, computing the intersection of the top  $G_{in}$  regions across rest and seven tasks. This is only an approximation since it does not conform fully to our definition of the global workspace, separate from the rest of the brain, which similar to a ‘club’ of functional hubs are characterized by a tendency to be more densely functionally connected among themselves than to other brain regions from where they receive integrative information. **b**, Nevertheless, this rough approximation shows a remarkable overlap with the Global Workspace derived from our formal definition in the FRIC algorithm applied to the DK80 parcellation (see Results and Fig. 4).



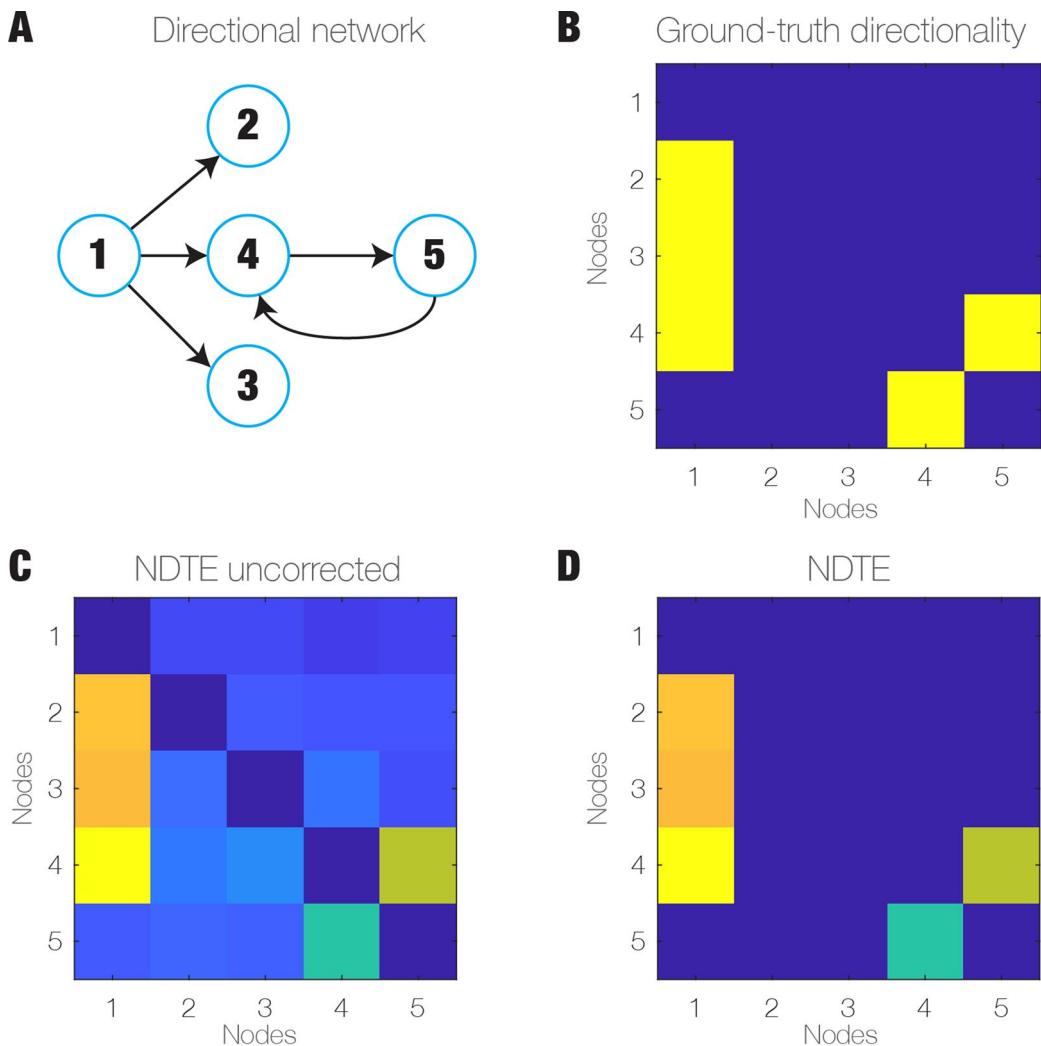
**Extended Data Fig. 6 | Comparing the anatomical rich club with the functional rich club (FRIC) for tasks and their intersection (Global Workspace).** **a**, Using the rich club algorithm, we computed the anatomical rich club on the structural connectivity matrix for DK80. The figure shows a rendering of the ten brain regions in the rich club: left and right precuneus, left and right thalamus, left and right superior frontal, left superior parietal, left insula and right isthmus cingulate cortices. **b**, A comparison with the global workspace (shown here as well in Fig. 4 and described in details in the results) shows that they overlap in the left precuneus and the right isthmus cingulate cortex. Given that the anatomical rich club cannot change over time with functional demands, the lack of further overlap is not surprising.



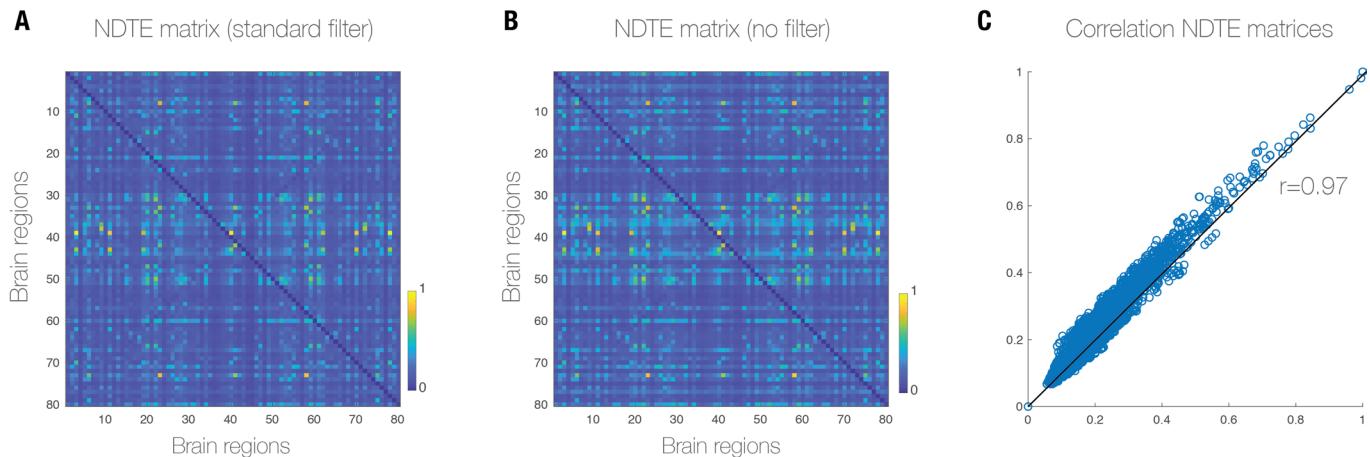
**Extended Data Fig. 7 | Estimation of structural connectivity (SC) matrices for DK80 using standard and special HCP datasets.** **a**, The 'standard SC matrix' was extract from the standard HCP multi-shell dMRI data from 985 HCP participants using state-of-the-art tractography methods (see Methods). **b**, The 'special SC matrix' was estimated using an even better state-of-the-art multi-shell dMRI protocol taking 89 minutes for each of 32 HCP participants at the MGH centre. Please note the increased number of inter-hemispheric connections (upper right and lower left quadrants of the special SC matrix compared to the standard SC matrix). **c**, The two SC matrices are highly correlated as shown in the scatterplot ( $r(3160)=0.85$ ,  $P < 0.001$ , 95% CI = 0.84–0.86).



**Extended Data Fig. 8 | Alternative measure of hierarchy.** Inspired by the neuroanatomical research by Markov and colleagues, we computed a similar measure of hierarchy as the fraction of feedforward and feedback (FF) organisation. We computed this for the seven tasks and for resting state (top row). As can be seen from the correlation matrices, this FF Hierarchy measure is similar to  $G_{out}$  for all seven tasks (mean correlation is 0.81 and the mean p-values are  $p < 0.001$ ). On the other hand FF Hierarchy is only non-significantly correlated with  $G_{in}$ , the integrative measure of incoming information flow (mean correlation is  $-0.02$  and the mean p-value is  $p = 0.243$ ).



**Extended Data Fig. 9 | Validation of NDTE framework.** Following the work of Seth, Chorley and Barrett<sup>1</sup>, we constructed a directional network with five nodes and simulated BOLD timeseries by using a Hopf model. **b**, This network is represented by the simple directional coupling  $5 \times 5$  matrix. The generative Hopf model was run 10 times with 1200 timepoints sampled at  $TR = 0.72$  s. We then validated the NDTE framework by applying this to the simulated data (using the same lag of 10 and 100 surrogates as in the analysis of the empirical data). **c**, The figure shows the recovered NDTE uncorrected matrix without using surrogates and aggregation of p-values. **d**, In contrast, the figure shows the results of using surrogates and aggregation of p-values and corrected for multiple comparisons. The NDTE framework is clearly able to recover the original couplings.



**Extended Data Fig. 10 | The estimation of the NDTE matrix does not depend on the filtering.** **a**, We computed the NDTE matrix in the DK80 parcellation from 100 unrelated HCP participants using a standard filter of 0.01–0.2 Hz on the underlying BOLD data. **b**, Similarly, we computed the NDTE matrix for the same participants with no filter. **c**, The correlation between the two NDTE matrices for the filtered and non-filtered version is almost identical ( $r(3160)=0.972$ ,  $P < 0.001$ , 95% CI = 0.971–0.973).

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection Standard MRI and MEG scanner software

Data analysis Standard imaging tools for preprocessing FSL (FMRIB Software Library), FreeSurfer, and the Connectome Workbench software.  
Custom-made MATLAB scripts were used for whole-brain modelling, which are freely available

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The code is freely available from github and multimodal neuroimaging data from the experiment are available on HCP website

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](http://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We used data of 1003 participants from the public available database from the Human Connectome Project (HCP) from the WU-Minn HCP Consortium.
Data exclusions	N/A
Replication	Findings from full dataset was replicated in 45 individuals with retest data
Randomization	N/A
Blinding	N/A

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input checked="" type="checkbox"/> MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	We used data of 1003 participants from the public available database from the Human Connectome Project (HCP) from the WU-Minn HCP Consortium.
Recruitment	The data set used for this investigation was selected from the March 2017 public data release from the Human Connectome Project (HCP) where we chose the full sample
Ethics oversight	HCP

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Magnetic resonance imaging

### Experimental design

Design type	Resting state and task design
Design specifications	1003 HCP participants. The HCP website ( <a href="http://www.humanconnectome.org/">http://www.humanconnectome.org/</a> ) provides the full details of participants, the acquisition and preprocessing of the data.
Behavioral performance measures	N/A

## Acquisition

Imaging type(s)	functional and diffusion MRI
Field strength	3T
Sequence & imaging parameters	The 1003 HCP participants were scanned on a 3-T connectome-Skyra scanner (Siemens). The HCP website ( <a href="http://www.humanconnectome.org/">http://www.humanconnectome.org/</a> ) provides the full details of participants, the acquisition and preprocessing of the data.
Area of acquisition	Whole brain
Diffusion MRI	<input checked="" type="checkbox"/> Used <input type="checkbox"/> Not used
Parameters	The HCP website ( <a href="http://www.humanconnectome.org/">http://www.humanconnectome.org/</a> ) provides the full details of parameters for the dMRI data.

## Preprocessing

Preprocessing software	Standardized methods using FSL (FMRIB Software Library), FreeSurfer, and the Connectome Workbench software
Normalization	FLIRT
Normalization template	CIFTI
Noise and artifact removal	The HCP website ( <a href="http://www.humanconnectome.org/">http://www.humanconnectome.org/</a> ) provides the full details of the preprocessing of the data.
Volume censoring	As per HCP pipeline

## Statistical modeling & inference

Model type and settings	Whole-brain model
Effect(s) tested	Spatiotemporal structure
Specify type of analysis:	<input checked="" type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input type="checkbox"/> Both
Statistic type for inference (See <a href="#">Eklund et al. 2016</a> )	N/A
Correction	N/A

## Models & analysis

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Functional and/or effective connectivity
<input checked="" type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input checked="" type="checkbox"/>	<input type="checkbox"/> Multivariate modeling or predictive analysis
Functional and/or effective connectivity	Pearson correlation