Toward Reliable Network Neuroscience for Mapping Individual Differences

Chao Jiang[1,2,5], **Richard Betzel**[3], **Ye He**[4], **and Xi-Nian Zuo**[1,2,5,6,7] ✉

[1]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China
[2]National Basic Science Data Center, Beijing 100190, China
[3]Department of Psychological and Brain Sciences, Indiana University, Bloomington, United States
[4]School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China
[5]Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
[6]IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China
[7]Key Laboratory of Brain and Education Sciences, Nanning Normal University, Nanning, Guangxi 530001, China

A rapidly emerging application of network neuroscience in neuroimaging studies has provided useful tools to understand individual differences in complex brain function. However, the variability of methodologies applied across studies - with respect to node definition, edge construction, and graph measurements-makes it difficult to directly compare findings and also challenging for end users to select the optimal strategies for mapping individual differences in brain networks. Here, we aim to provide a benchmark for best practices by systematically comparing the reliability of human brain network measurements of individual differences under different analytical strategies using the test-retest design of the resting-state functional magnetic resonance imaging from the Human Connectome Project. The results uncovered four essential principles to guide reliable network neuroscience of individual differences: 1) use a whole brain parcellation to define network nodes, including subcortical and cerebellar regions, 2) construct functional connectome using spontaneous brain activity in multiple slow bands, 3) optimize topological economy of networks at individual level, 4) characterise information flow with metrics of integration and segregation.

reliability | network | connectome | resting-state fMRI | open science

Correspondence: *xinian.zuo@bnu.edu.cn, zuoxn@psych.ac.cn*

Over the past two decades, network neuroscience has helped transform the field of neuroscience, providing a quantitative framework for modeling brains as graph (or networks) composed of nodes (brain regions) and edges (their connections) (1). The organization and topology of brain networks can be characterized by a growing suite of graph-theoretic tools, including measurements of their global efficiency, nodal centrality, clustering, small-word topology, and many other graph metrics (2–4). In parallel, resting-state fMRI (rfMRI) has opened up new avenues towards understanding the human brain function (5). In conjunction with network neuroscience, rfMRI has led to the emergence of a multidisciplinary field, functional connectomics (6–8), in which the brain's intrinsic, interregional connectivity is estimated from rfMRI recordings. Functional connectomics has been widely used to investigate the system-level organization of the human brain (9) and its relationships with individual differences (10) in developmental (11), socio-cultural (12) and clinical conditions (13).

An important topic in functional connectomics and, indeed, any scientific discipline, is the notion of measurement reliability. In general, reliability characterises a proportion of measurement variability between different subjects relative to the overall variability including both between- and within-subject, i.e. random, components (14), and is commonly used to assess the consistency or agreement between measurements. However, it can also serve as a measure of discriminablity. For example, if a measurement can sufficiently capture individual characteristics, its reliability will be higher than measurements that underestimate between-subject variability. Thus, high reliability is essential for any measurement to better differentiate a group of individuals, i.e., inter-individual differences (15). Recent studies have demonstrated that the reliability of measurements is equivalent to the fingerprint and distinguishability of the measurement under the Gaussian distribution (16) while it has well-established statistical theory and applications to psychology, i.e., psychometric theory (17) and medicine, i.e., diagnosis theory (18). Reliability also provides an upper bound of the measurement validity (4, 15), which cannot be readily quantified as the reliability (15). Therefore, high levels of reliability is the first and most basic requirement for quantifying individual differences in functional connectomics. Accordingly, the optimization of measurement reliability can help guide processing and analysis pipelines for neurodevelopmental (19) and clinical applications (20).

Previous studies have demonstrated that many measurements made on networks estimated from rfMRI have limited reliability (21, 22). These low levels of reliability could be an indication of failure in handling individual variability at different levels (23). In particular, experimental design and processing decisions related to scan duration, determining frequency range, and regressing global signal have impacts on rfMRI measurements and thus their reliability (22, 24). Although less focused on reliability, existing studies also revealed that their findings are influenced by choices of parcellation templates (25), edge construction and definition, and choice of graph metrics (26). How these decisions affect the reliability of functional connectomics deserves further investigation. These analytical choices have been implemented in different software packages (27) but can vary from one package to another in terms of their parameterization. Beyond

limited examinations on reliability (28–30), a systematic investigation into the reliability of network neuroscience with rfMRI is warranted to guide software use and future functional connectomics analyses.

In this paper, we conducted a systematic investigation of reliability for network neuroscience based upon the test-retest rfMRI data from the Human Connectome Project (HCP) (31). Note that the HCP imaging acquisition settings and data preprocessing have integrated various strategies to optimize the measurement reliability (21, 32). We thus analyzed the minimally pre-processed HCP rfMRI data and focused our work on four key post-processing stages: (1) node definition, (2) edge construction, (3) network measurement, and (4) reliability assessments. In the end, we propose a set of principles to guide researchers in performing reliable network neuroscience with rfMRI, advancing the field-standard call for the best practices in clinical network neuroscience (23). Toward an open network neuroscience, we shared all the codes and reliability maps by building an online platform (ADD URL) for sharing the reliability data and computation.

Results

A typical analysis pipeline in functional connectomics includes steps for node definition (parcellations) and edge construction (frequency bands, connectivity estimation and filtering schemes) (Fig. 1a). To determine an optimal pipeline, our aim is to combine the most reliable strategies across different parts of the analysis by comparing the reliability of derived global network metrics. The HCP test-retest data were employed for reliability evaluation (Fig. 1b) using the intraclass correlation (ICC) statistics on the measurement reliability with five levels (34): $0 < \mathrm{ICC} \leq 0.2$ (**slight**); $0.2 < \mathrm{ICC} \leq 0.4$ (**fair**); $0.4 < \mathrm{ICC} \leq 0.6$ (**moderate**); $0.6 < \mathrm{ICC} \leq 0.8$ (**substantial**); and $0.8 < \mathrm{ICC} < 1.0$ (**almost perfect**). Our analyses produce massive amounts of reliability statistics: 524,160 ICCs. In this section we first present overall reliability assessments associated with the various analytic strategies as well as their impact on between- and within-subject variability (Fig. 1c). We then determine the optimized pipelines based on the highest reliability measurements, while documenting the derived both global and local network metrics and both their reliability and variability at an individual level. Based upon these results, we built the open resources for reliable functional connectomics, including all the codes, reliability matrices and computation *via* an online platform for resource sharing (ADD URL).

Whole brain networks are more reliable than cortical networks. Elements derived from a brain parcellation (i.e., parcel) define the network nodes. Here, we evaluated reliability based on 30 different parcellation choices (Fig. 2a, see more details of these parcellations in **Materials and Methods**). In the following parts of the paper, we name a parcellation as '*ParcAbbr-NumberOfParcels*' (e.g., LGP-100 or its whole-brain version wbLGP-458).

We found significant differences in ICC distributions across the 30 parcellation choices (Fig. 2b, Friedman rank sum test: $\chi^2 = 28955.83, df = 29, p < 2.2 \times 10^{-16}$, effect size $W_{Kendall} = 0.425$). The mean ICCs range from slight (wbGICAP-15) to substantial (wbLGP-458). Given a particular parcellation and definition of nodes, we illustrate the density distribution of its ICCs under all other strategies (edge definition and metric derivation). Notably, whole-brain parcellations (with the exception of wbGICAPs) yield higher measurement reliability than parcellations of cerebral cortex on their own (the effect sizes r(rank-biserial correlations)$>$ 0.65). This improvement in reliability seems not simply a biproduct of having more parcels. We chose the parcellations in which the number of parcels ($400 \leq n \leq 1000$) almost overlapped between the cortex and the whole brain, and found no correlation between the number of parcels and the median ICCs ($r = -0.11, p = 0.7$). We report the mean ICC and the number of almost perfect (noap) ICCs (≥ 0.8) as the descriptive statistics for the density distributions. The wbLGP-458 (mean ICC: 0.671; noap ICC: 519), wbLGP-558 (mean ICC: 0.671; noap ICC: 540) and The wbBNP-568 (mean ICC: 0.664; noap ICC: 511) are the three most reliable choices (see more details of the post-hoc Wilcoxon signed rank test in Table S5). Among the cortical parcellations, the LGP-500 (mean ICC: 0.362; noap ICC: 0), LGP-400 (mean ICC: 0.342; noap ICC: 0) and LGP-600 (mean ICC: 0.340; noap ICC: 0) are the three most reliable choices (Table S5). Performances of the ICA-based parcellation choices ranked below the cortical parcellation choices depending on their sizes: wbGICAP-300 (mean ICC: 0.241; noap ICC: 0) and wbGICAP-200 (mean ICC: 0.214; noap ICC: 0) while other wbGICAP choices showed the lowest ICCs.

To better understand the effect of introducing 358 subcortical parcels into the cortical parcellations, we decomposed the reliability changes into a two-dimensional representation of changes of individual variability (Fig. 2c,d). This idea was motivated by the analysis of reliability derived with individual variability (14, 15) as in Fig. 1c. For each ICC under a given parcellation choice, we calculated the related between-subject variability V_b and within-subject variability V_w. Changes in the individual variability associated with the reliability improvements from cortical to whole-brain pipelines were plotted along with ΔV_b and ΔV_w as arrows. These arrows are distributed across the three quadrants (quadI: 0.94%; quadII: 59.99%; quadIII: 39.07%). We noticed that most of these arrows were distributed into the optimal quadrant where the improvements of test-retest reliability by the whole-brain parcellation choices largely attributing to the increases of between-subject variability and decreases of within-subject variability. The decreases of both between-subject and within-subject variability may also strengthen the measurement reliability (the suboptimal quadIII in Fig. 2).

Spontaneous brain activity portrays more reliable networks in higher slow bands. Brain oscillations are hierarchically organized, and their frequency bands were theoretically driven by the natural logarithm linear law (33, 35). By analogy, rfMRI oscillations can, similarly, be partitioned into distinct frequency bands. Advanced by the fast imaging protocols (TR = 720ms), HCP test-retest data allows us to

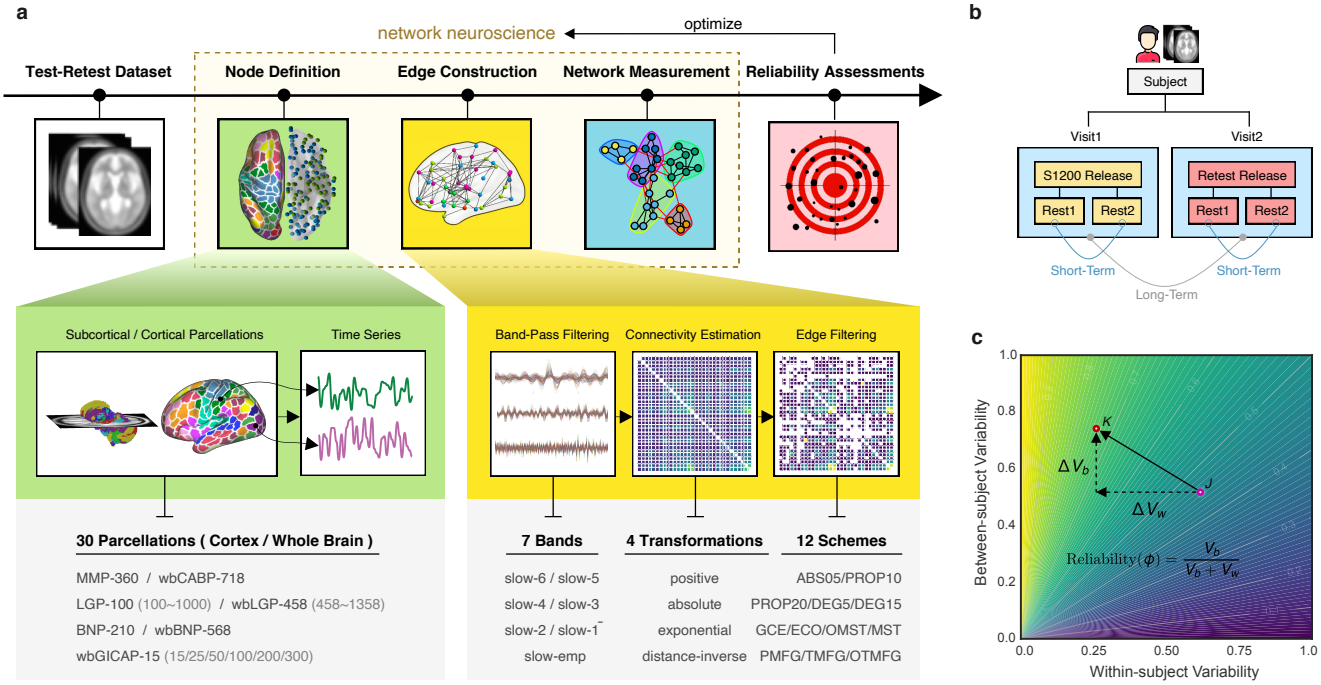

Fig. 1. Analytical pipelines for reliable functional connectomics. a) There are five stages during our analyses: (1) test-retest dataset (white box) downloaded from HCP website, (2) node definition (green box) defining nodes using a set of brain areas of 30 different partitions of the human brain, (3) edge construction (yellow box) estimating individual correlation matrices using the six frequency bands (slow 1-6) from Buzsáki's theoretical framework on the brain oscillations (33) as well as the widely used empirical frequency band (Slow-emp) (5) and transferring these matrices into adjacency matrices using $7 \times 4 \times 12$ different strategies on edge construction including band-pass filtering, connectivity estimation and edge filtering, (4) network analysis (blue box) systematically calculating various brain graph metrics on measurements of information flow, and (5) reliability assessment (red box) evaluating test-retest reliability with massive linear mixed models. b) The test-retest data shared multimodal MRI datasets of 46 subjects in the HCP S1200 release and the HCP Retest release. Each subject underwent the first four test scans on two days (two scans per day: REST1 and REST2) and return several months later to finish the four retest scans on another two days. This was designed for evaluation of both the short-term (REST1 vs. REST2) and the long-term (VISIT1 vs. VISIT2) reliability. c) Measurement reliability refers to the inter-individual or between-subject variability V_b relative to the intra-individual or within-subject variability V_w. Variability of both between-subject (V_b) and within-subject (V_w) are normalized into between 0 and 1 by the total sample variances. Their changes (ΔV_b and ΔV_w) introduce a reliability gradient as represented by the vector (the black arrow). The length of the arrow reflects the amplitude of reliability changes when the reliability assessment from one choice (pink circle, J) to another choice (red circle, K). Further, the arrow's direction ($\mathbf{JK}$) indicates the sources of this reliability change. Here the reliability becomes from moderate to substantial level with increases of between-subject variability ($\Delta V_b > 0$) and decreases of within-subject variability ($\Delta V_w < 0$).

obtain more oscillation classes than traditional rfMRI acquisitions (typical sampling frequency of TR = 2s). We incorporate Buzsáki's framework (33, 36) with the HCP dataset by using the DREAM toolbox (37) in the Connectome Computation System (27). It decomposed the time series into the six slow bands (Fig. 3a):

- **slow-6** (0.0069-0.0116 Hz)
- **slow-5** (0.0116-0.0301 Hz)
- **slow-4** (0.0301-0.0822 Hz)
- **slow-3** (0.0822-0.2234 Hz)
- **slow-2** (0.2234-0.6065 Hz)
- **slow-1**$^-$ (0.6065-0.6944 Hz)

We noticed that, due to the limited sampling rate (TR), this **slow-1**$^-$ only covers a small part of the full **slow-1** band (0.6065-1.6487 Hz) – we indicate this above. We also included the frequency band, **slow-emp** (0.01-0.08 Hz) for the sake of comparison, as it is covers a range commonly used in fMRI studies. A significant effect on order ($\chi^2 = 9871.291, df = 6, p < 2.2 \times 10^{-16}, W_{Kendall} = 0.163$) across the frequency bands was revealed based on the ICC density distributions (Fig. 3b): slow-2, slow-1$^-$, slow-3, slow-emp, slow-4, slow-5, slow-6. Post-hoc paired tests indicated that any pairs of neighbouring bands are significantly different from one another (for more details, see Table S10-12), with measurement reliability increasing with faster frequency bands. Note, however, that slow-1$^-$ (mean ICC: 0.498) did not fit into this trend, possibly due to its limited coverage of the full band. But remarkably, slow-1$^-$ exhibited the largest number of almost prefect ICCs for potential reliability (noap ICC: 1746, for more details, see Figure S9). Slow-emp (mean ICC: 0.467; noap ICC: 434) contains overlapping frequencies with both slow-4 (mean ICC: 0.450; noap ICC: 441) and slow-5 (mean ICC: 0.409; noap ICC: 285), and higher ICCs than the two bands but the effect sizes are small to moderate (slow-emp vs. slow-4: 0.193; slow-emp vs. slow-5: 0.485). Slow-6 is the choice with the lowest ICCs (mean ICC: 0.278; noap ICC: 154) compared to other bands (large effect sizes: $r > 0.582$).

To visualize variation in reliability across frequency bands, we plotted a trajectory tracing reliability flow along the five full (slow-6 to 2) bands in the reliability plane, whose axes correspond to between- *versus* within-subject variability (Fig. 3c). As expected, this nonlinear trajectory contains two stages of almost linear changes of the network measure-

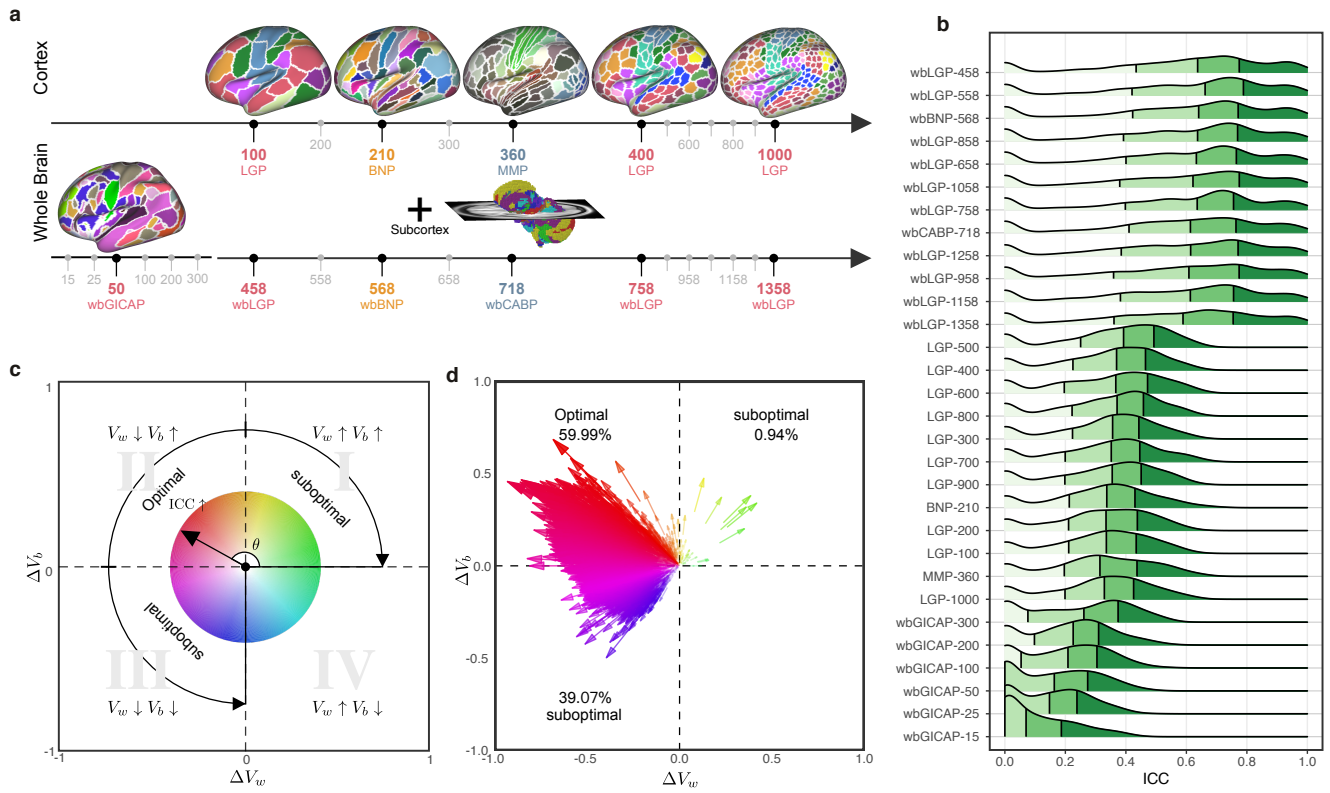

Fig. 2. Parcellation choices impact measurement reliability and individual variability. a) Node definitions are derived from different resolutions of spatial parcellation on the human cortex and whole brain (see more details of these name abbreviations in **Materials and Methods**). b) Density plots are visualized for distributions of the ICCs under the various parcellation choices on node definition. These density distributions are ranked from top to bottom according to decreases of the mean ICCs while the four colors depict the four quantiles. c) Reliability gradient between any one whole-brain parcellation choice and its corresponding cortical parcellation choice is decomposed into the axis of changes of the between-subject variability (ΔV_b) and the axis of changes of the within-subject variability (ΔV_w). This gradient can be represented as an vector, which is the black arrow from the origin with an angle θ with the $x-$axis while the color encodes this angle and the transparency or the length reflects the magnitude of the degree of ICC improvement. According to the anatomy of reliability, the optimal space is in the second quadrant (quadII) while the first and third quadrant (quadI and quadIII) are suboptimal for reliability. d) Improvements of reliability by the whole-brain node definition pipelines are represented as gradient arrows in the plane of individual variability.

ment reliability from slow to fast oscillations: whole brain versus cortex. In each case, the reliability improvements attribute to both increases of between-subject variability and decreases of within-subject variability while the improvements of whole-brain network measurement reliability were largely driven by the increased variability between subjects.

Topological economics individualize highly reliable functional brain networks. Estimating functional connections can be highly challenging due to the absence of a 'ground truth' human functional connectome. To provide a reliable way of building candidate edges of the connections, we sampled the 12 schemes on graph edge filtering (Fig. 4a), which turn a fully connected matrix into a sparse graphical representation of the corresponding brain network. These schemes can be categorized into two classes: threshold-based *versus* topology-based schemes. Absolute weight thresholding $(\text{ABS})_{05}$, proportional thresholding $(\text{PROP})_{10}$, PROP_{20}, degree thresholding $(\text{DEG})_5$, DEG_{15}, overall efficiency cost optimization (ECO) and global cost efficiency optimization (GCE) commonly employ an threshold for filtering edges with higher strengths than a cut-off value. These schemes are widely used in network neuroscience and ignore the intrinsic topological structure of the entire brain network (e.g, leading to multiple connected components or isolated nodes). In contrast, topology-based schemes such as minimum spanning tree (MST), orthogonal MST (OMST), planar maximally filtered graph (PMFG) and triangulated maximally filtered graph (TMFG) come from other scientific disciplines and are optimized based on the entire network topology (39–42). To combine both the TMFG's efficiency and OMST's accuracy, we proposed the orthogonal TMF graph (OTMFG). All the schemes are plotted in the plane of cost *versus* global-cost efficiency (43) to better visualize the economical properties of the derived networks (Fig. 4b). These plots are fitted into the topographic (contour) maps where the local maxima for each filtering choice is labeled as a circle. The human brain networks achieve higher global efficiency with lower cost using topology-based schemes compared to threshold-based schemes, suggesting increasingly optimal economics.

Significant differences in test-retest reliability were detectable across these 12 edge-filtering schemes ($\chi^2 = 8749.223, df = 11, p < 2.2 \times 10^{-16}, W_{Kendall} = 0.135$, see Fig. 4c). Among the topology-based schemes, OMST (mean ICC: 0.538; noap ICC: 765), OTMFG (mean ICC: 0.531; noap ICC: 781) and TMFG (mean ICC: 0.502; noap ICC: 767) were the three most reliable choices. They showed significantly greater reliability than the three most reliable threshold-based, respectively: PROP_{20} (mean ICC: 0.527; noap ICC: 632), PROP_{10} (mean ICC: 0.488; noap ICC: 445)

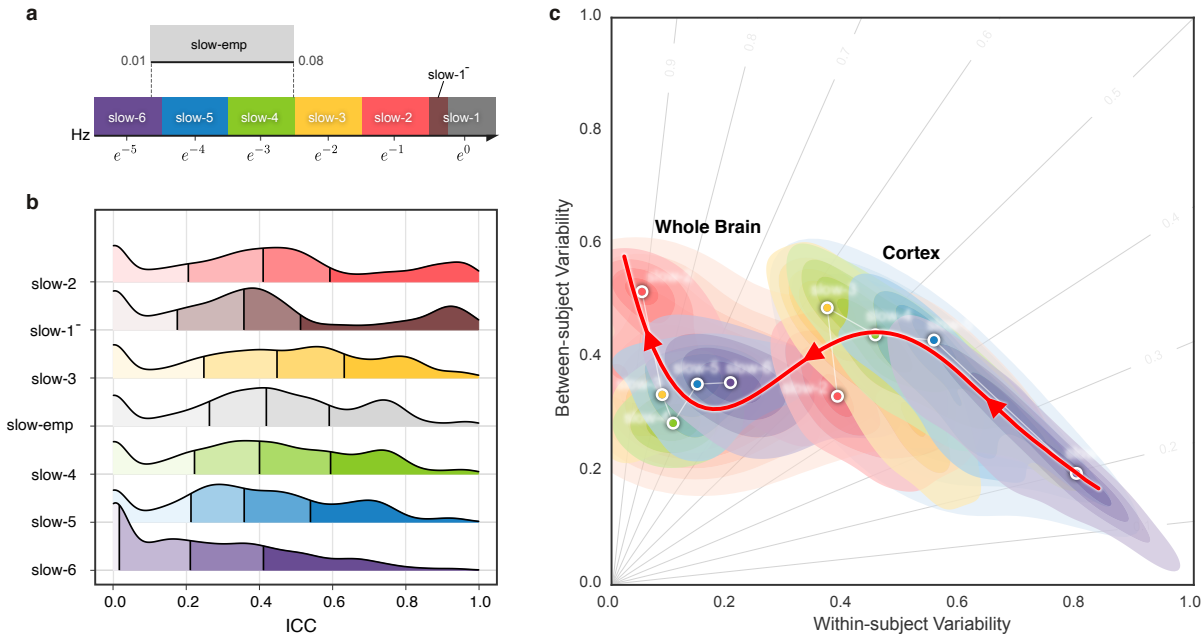

Fig. 3. Reliability gradient across the slow bands and changes of related individual variability. a) Classes of frequency bands for slow oscillations derived from the natural logarithm linear law (33, 35). b) Density plots are visualized for the ICC distributions under the various frequency bands in edge construction. These density distributions are ranked from top to bottom according to decreases of the mean ICCs while the vertical lines depict the four quartiles. c) Network measurements are projected onto the reliability anatomy plane coordinated by both between- and within-subject variability. These dot plots are fitted into the topographic (contour) maps where the local maxima for each band is labeled as a circle. The red line shows the entire modeled trajectory tracing the reliability flow along slow-to-fast oscillations in cortex and whole brain.

and GCE (mean ICC: 0.473; noap ICC: 352). Mean reliability of MST are slight to fair (mean ICC: 0.267) but its number of almost perfect reliability (noap ICC:362) is still higher than all threshold-based schemes except PROP_{10} and PROP_{20} (see Figure S23).

Network measurements are labeled based on topology and threshold groups and projected onto the reliability anatomy plane, whose axes represent between- and within-subject variability (Fig. 4d). The contour maps are reconstructed for each scheme based upon the individual variability of all the related network measurements. The topology-based methods (red) showed overall higher ICCs than the threshold-based methods (blue), improvements that could be attributed to increases in between-subject variability and decreases of within-subject variability. These observations are consistent between cortex and whole brain networks while topology-based whole brain network are almost perfectly reliable (meaning almost perfect reliability, i.e., $\text{ICC} \geq 0.8$).

We also explored connection transformation and edge weights, two factors included in edge filtering, the choices of connectivity transformation and weighing edges, in terms of their measurement reliability. Positive (Eq.pos) (mean ICC: 0.453; noap ICC: 1,031) and exponential (Eq.exp) transformation (mean ICC: 0.451; noap ICC: 1,855) were the two most reliable transformation choices. Comparing to the positive and absolute (Eq.abs) (mean ICC: 0.447; noap ICC: 1,050) transformation, the exponential and distance-inverse (Eq.div) (mean ICC: 0.442; noap ICC: 1,031) transformation show larger number of almost perfect ICCs (see Table S13-19). Weighted graphs are also more reliable than the binary graphs while the normalized weighted graphs demonstrated the highest ICCs. Such improvements reflect both the increased between-subject variability and decreased within-subject variability (Figure S1).

Network integration and segregation can serve reliable metrics of information flow. The previous big data analysis suggests that the optimally reliable pipeline should: 1) define network nodes using a whole-brain parcellation, 2) filter the time series with higher frequency bands, 3) transform the connectivity using positive transformation, 4) construct network edges using individualized methods and normalized weights. Using the optimal pipelines, we evaluated the reliability levels of various metrics from network neuroscience and their differences across individuals. Focusing on the optimized pipeline with the highest ICCs of the various choices (wbLGP-458, slow-2, pos, OMST), we reported test-retest reliability of the measurements as well as their corresponding individual variability. In Fig. 5a, we found that the global network measurements of information segregation and integration are at the level of almost perfect reliability except for the modularity Q (ICC=0.46, 95% CI = [0.252,0.625]). These high-level ICCs are derived with large between-subject variability and small within-subject variability (Fig. 5b). These findings are reproducible across the other two parcellation choices (wbCABP-718, wbBNP-458).

Similar to the global metrics, shortest path length L_p and nodal efficiency E_{nodal} exhibited the highest ICCs (almost perfect test-retest reliability) while ICCs of other nodal metrics remained less than 0.6. To visualize node-level net-

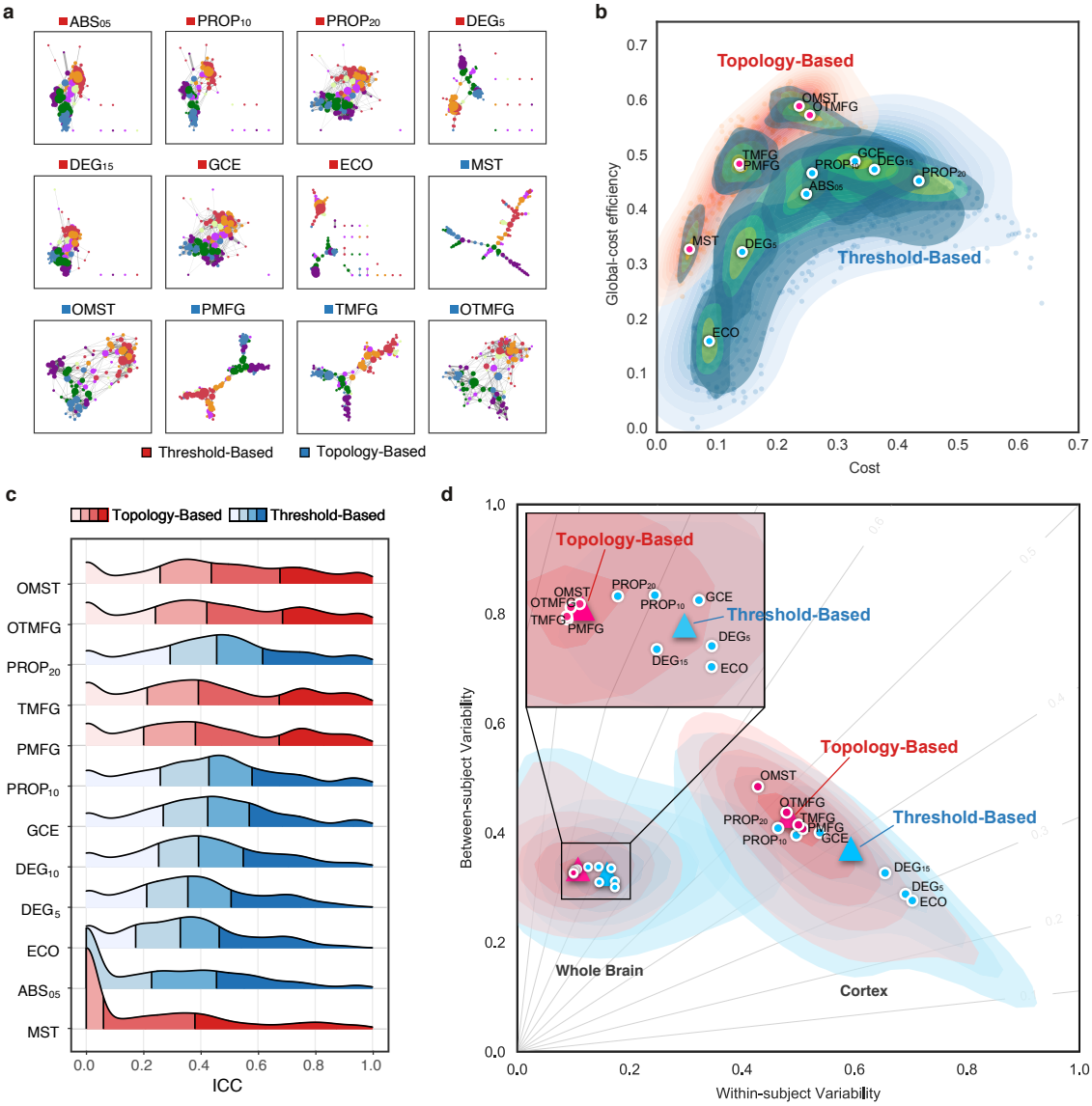

Fig. 4. Edge filtering schemes and their networking performance. (a) Twelve schemes of filtering edge are applied to an individual connectivity matrix, resulting in the 12 brain networks with their nodes colored as the Yeo2011-7Networks (38). (b) Global cost efficiency are plotted against network wiring costs of all the brain networks derived with the 12 edge filtering schemes from all the individual rfMRI scans. Red dots represent the topology-based while blue dots are for threshold-based networks. These dot plots are fitted into the topographic (contour) maps where the local maxima for each filtering choice is labeled as a circle. (c) Density plots are for ICC distributions under various the 12 edge filtering schemes. These density distributions are ranked from top to bottom according to decreases of the mean ICCs while the two colors depict the topology-based and threshold-based schemes. Four quartiles were indicated by vertical lines. (d) Network measurements are projected onto the reliability anatomy plane coordinated by both between- and within-subject variability. Red dots represent the topology-based while blue dots are for threshold-based networks. The topographic (contour) maps fit the dots and label the local maxima as a circle for each scheme and the global maxima as a triangle for the topology and threshold groups, respectively.

work metrics, we reported results derived from the wbCABP-718 choice. To improve spatial contrasts of reliability, we ranked the parcels according to their ICCs and visualized the ranks in Fig. 5c. Most nodal metrics are more reliable across the 360 cortical areas than the 358 subcortical areas (Wilcoxon tests: all p-values less than 0.001, corrected for multiple comparisons). However, L_p, E_{nodal} and B_c exhibited higher across subcortical areas than cortical areas (corrected $p < 0.001$). Across the human cerebral cortex, the right hemispheric areas demonstrated more reliable C_p (corrected $p < 0.0036$) than the left hemispheric areas. Interesting patterns of the reliability gradient are also observable along large-scale anatomical directions (dorsal>ventral, posterior>anterior) across the nodal metrics of information segregation and centrality. These spatial configuration profiles on the reliability reflected their correspondence on inter-individual variability of these metrics, characterising the network information flow through the slow-2 band.

Building an open resource for reliable network neuroscience. The results presented here represent huge costs in terms of computational resources (more than 1,728,000 core-hours on **CNGrid**, supported by Chinese Academy of Sciences (http://cscgrid.cas.cn)). Derivations of the ICCs and their linear mixed models were implemented in **R** ad **Python**. As our practice in open science, we have

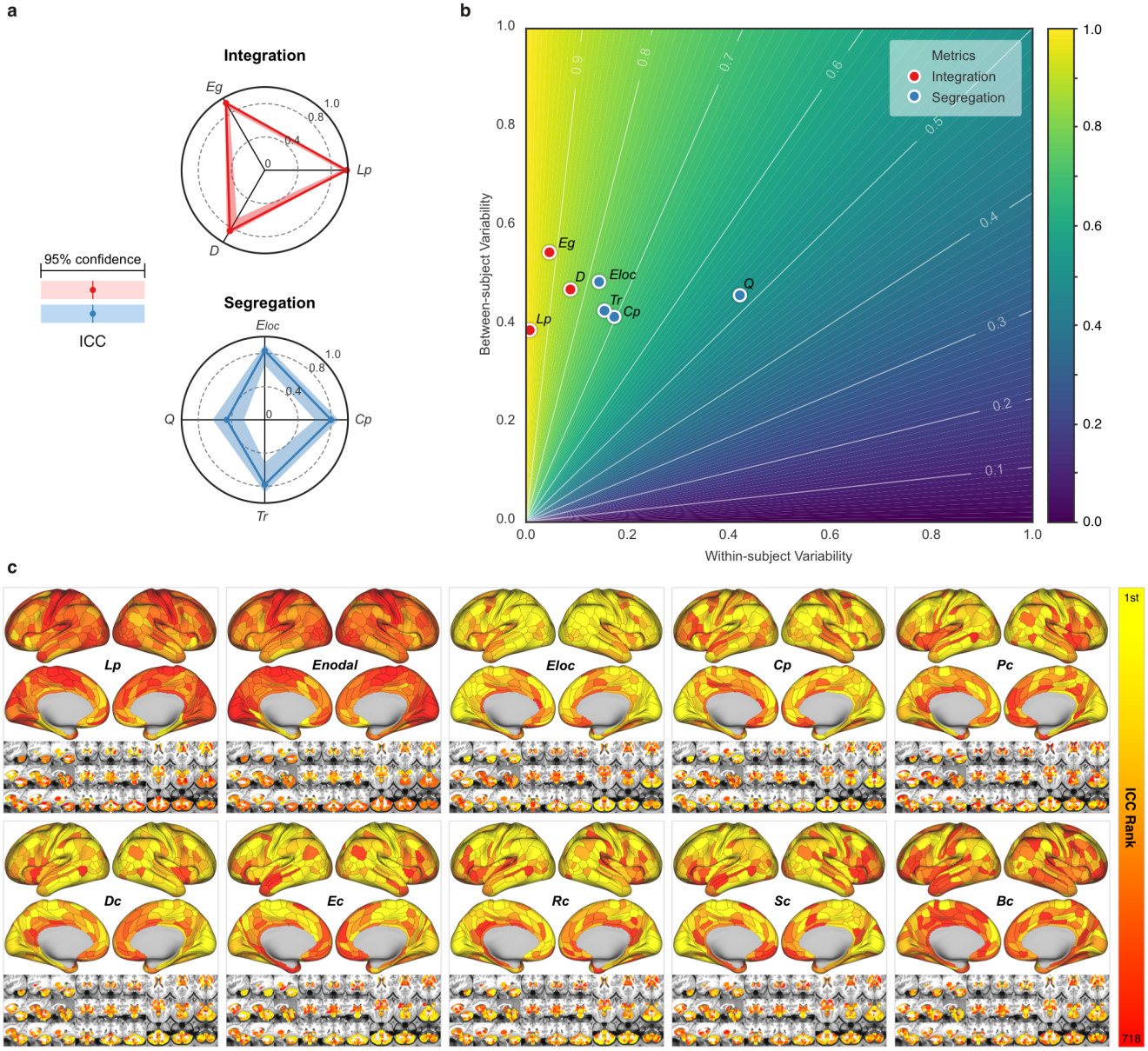

Fig. 5. Measurement reliability and variability of global/nodal network metrics under the optimized pipeline. (a) Spider plots are visualized for ICCs (test-retest) with the 95% confidence intervals (CIs, shadow bands) for the metrics of network integration, segregation, centrality and resilience. The associated symbols of the metrics can be found in Appendix Table 1. (b) The reliability anatomy was plotted as a function of between-subject variability (V_b) and within-subject variability (V_w). (c) Ranks of ICCs across the 360 cortical parcels and the 358 subcortical parcels in the optimal pipeline (wbCABP-718, slow-2, pos, OMST) are depicted. Ten nodal metrics are assessed including average shortest path length L_p, nodal efficiency E_{nodal}, local efficiency E_{local}, clustering coefficient C_p, pagerank centrality P_c, degree centrality D_c, eigenvector centrality E_c, resolvent centrality R_c, subgraph centrality S_c and betweeness centrality B_c.

started to provide an online platform on the reliability assessments (ADD URL). Detailed documentation can be found at the online website (ADD URL). The big reliability data were designed into an online database for providing the community a resource to search reliable choices and help the final decision-making. More details of the database use can be found from the online documentation (ADD URL). We shared all the codes and figures as well as other resources via (ADD URL) for reliable functional network neuroscience.

Discussions

This study examined the series of processing and analysis decisions in constructing graphical representations of brains. The focus, here, was on identifying the pipeline that generated reliable, individualized networks and network metrics. The results of our study suggest that to derive reliable global network metrics with higher inter-individual variances and lower inner-individual variances, one should use whole-brain parcellations to define network nodes, focus on higher frequencies in the slow band for time-series filtering to derive the connectivity, and use topology-based methods for edge filtering to construct sparse graphs. Regarding network met-

rics, multi-level or multi-modal metrics appear more reliable than single-level or single-model metrics. Derive reliable measurements is critical in network neuroscience, especially for translating network neuroscience into clinical practice, which requires precise and specific biomarkers. Based on these results, we provide four principles of reliable functional connectomics which we discuss further in this section.

Principle I: Use a whole brain parcellation to define network nodes. The basic unit of a graph is the node. However, variability across brain parcellations can yield dissimilar graphs, distorting network metrics and making it difficult to compare findings across studies(25, 44, 45) (46). In many clinical applications (13, 20), researchers aim to identify disease-specific connectivity profiles of the whole brain, including cortical and subcortical structures, as well as cerebellum. A recent review has raised the concern that many studies have focused on restricted sets of nodes, e.g. cortex only, called a field standard for the best practices in clinical network neuroscience (23), which requires almost perfectly reliable measurements (14, 47). Our meta-reliability assessments revealed high reliability of measurements made involving functional brain networks can be achieved, namely through the inclusion of high-resolution subcortical nodes. This provides strong evidences that the whole-brain node use should be part of the standard analysis pipeline for network neuroscience applications. These improvements of reliability can be attributed to increases in between-subject variability coupled with reductions in within-subject variability relative to networks of cortical regions alone. One possible neuroanatomical explanation is that distant areas of cerebral cortex are interconnected by the basal ganglia and thalamus (48) while also communicating with different regions of the cerebellum *via* polysynaptic circuits (49, 50), forming an integrated connectome. These subcortical structures have been suggested to play a role in both primary (e.g., motor) as well as higher-order function (e.g., learning and memory (51)). Studies using rfMRI have delineated the resting-state functional connectivity (RSFC) maps between these subcortical structures and cortical networks of both primary and high-order functions (52–54). A recent work revealed that inter-individual variance in cerebellar RSFC networks exceeds that of cortex (55). Meanwhile, these RSFC maps are highly individualized and stable within individuals (56–58), indicating that they possess reliable characteristics. In line with our observations, we argue that inclusion of the subcortical structures as network nodes can enhance the between-subject variability and stabilize the within-subject variability by providing a more comprehensive measurements on the entirety of the brain connectivity. Larger between-subject variability implies that the associated measurements are more recognizable between different subjects, leading to improved subject discrimination, a finding that has been demonstrated (59, 60).

Principle II: Generate functional networks using spontaneous brain activity in multiple slow bands. It has been a common practice in RSFC research area to estimate the RSFC profile based on the low-frequency (0.01 - 0.1 Hz or 0.01 - 0.08 Hz) fMRI time series (5). However, the test-retest reliability of measurements made based on this frequency band has been limited, with ICCs less than 0.4 (see (21, 22) for systematic reviews). Other applications, however, have advocated adopting a multi-frequency perspective to examine the amplitude of brain activity at rest (61) and its network properties (62). This approach has been spurred along by recent advances in multi-banded acquisitions and fast imaging protocols, offering fMRI studies a way to examine resting-state brain activity at relatively higher frequencies that may contain neurobiologically meaningful signals (37, 63). Our study provides strong evidence of highly reliable signals across higher slow-frequency bands, which are derived with the hierarchical frequency band theory of neuronal oscillation system (33). Specifically, a spectrum of reliability increases was evident from slow bands to fast bands. This reflects greater variability of the network measurements between subjects and less measurement variability within subject between the higher and lower bands of the slow frequencies. In theory, each frequency band has an independent role in supporting brain function. Lower frequency bands are thought to support more general or global computation with long-distance connections to integrate specific or local computation, which are driven by higher slow bands based on short-distance connections (35). Our findings of high reliability (inter-individual differences) are perfectly consistent with this theory from a perspective of individual variability. Previous findings have found that high-order associative (e.g., default mode and cognitive control) networks are more reliable than the primary (e.g., somatomotor and visual) networks (15, 21, 22). Our findings offer a novel frequency-based perspective on these network-level individual differences.

Principle III: Optimize topological economy to construct network connections at individual level. There is no gold standard on for human functional connectomes, leading to plurality of approaches for inferring and constructing brain network connections. Threshold-based methods focus on the absolute strength of connectivity, retaining connections that are above some user-defined threshold and oftentimes involve applying the same threshold to all subjects. Although this approach mitigate potential biases in network metrics associated with differences in network density, it may inadvertently also lead to decreased variability between subjects. This is supported by our result finding that threshold-based method yield low reliability of network measurements. On the other hand, the human brain is a complex network that is also near-optimal in terms of connectional economy, balancing tradeoffs of cost with functionality (64). In line with this view, certain classes of topology-based methods for connection definition may hold promise for individualized network construction. Specifically, each individual brain optimizes its economic wiring in terms of cost and efficiency, reaching a trade-off between minimizing costs and allowing the emergence of adaptive topology. Our results demonstrate that such highly individualized functional connectomes generated by the topology-based methods are more reliable than those by the threshold-methods. This reflects the increases

of individual differences in functional connectomes attributing to the optimal wiring economics at individual level. The topological optimization also brings other benefits such as ensuring that a graph forms a single connected component and preserving weak connections. Indeed ,there is increasing evidence supporting the hypothesis that weak connections are neurobiologically meaningful and explain individual differences in mind, behavior and demographics as well as disorders (65–67). Weak connections in a graph may be consistent across datasets and reproducible within the same individual over multiple scan sessions and therefore be reliable. Weak connections might also play non-trivial roles in transformed versions of the original brain network, e.g. so-called "edge-based functional connectivity" (68). Among these topology-based methods, MST is the simplest and promising filtering method if computational efficiency is the priority. MST can obtain a graph with the same number of nodes and edges, and it is not sensitive to scaling effects, because its structure only depends on the order rather than the absolute values of the edges (69). Although MST loses some local network measurements due to the limited number of edges, it has some other unique metrics that can be calculated (e.g., leaf fraction, tree hierarchy). A better alternative might be TMFG which computationally very efficient and statistically robust, while the OMST and OTMFG are the most reliable choices given priority to large individual differences.

Principle IV: Characterise information flow with network integration and segregation metrics. Functional connectomes reflect the outcome of communication processes and information flows between pairs of brain regions. How information and other signals propagate between pairs of brain regions can be assayed using network neuroscientific metrics and is essential to understanding normative connectome function and its variation in clinical settings (70). While the ground truth functional connectome is unknown (and may not exist (71)), network models can help validate the imaging-based reconstructions of human functional connectomes (1). From a perspective of individual differences, reliable functional connectomics is the basis of achieving valid measurements of the individual differences in functional connectomics (15). Our findings indicated that both the brain network segregation and integration could be reliably measured with functional connectomics using rfMRI by the optimized pipelines. At the global level, measures of information integration, e.g. characteristic path length and efficiency, were more reliable than those of information segregation, e.g. modularity and clustering coefficient. Our results also revealed that measures of integration were more stable across different scan sessions (i.e., the test-retest) for an individual subject than the segregation measurements while the inter-individual variability are measured at the similar level for both integration and segregation metrics. At nodal level, mapping reliability of the network measurements revealed interesting spatial patterns. Specifically, we found that cortical areas were generally associated with more reliable local measurements compared to subcortical areas. This may reflect different functional roles for human cortex and subcortex. For example, the differences in reliability of path-based metrics might reflect the fact that there are more within-community paths in cortex while between-community paths are more common in subcortex. Beyond this cortical-subcortical gradient, reliability of the nodal information flow also fit the left-right asymmetry and dorsal-ventral as well as posterior-anterior gradient, implying the potential validity of individual differences in information flow attributing to evolutionary, genetic and anatomical factors (72–75). To facilitate the utility of reliable network integration and segregation metrics in functional connectomics, we integrated all the reliability resources into an online platform (ADD URL) for reliability queries on specific metrics of information flow integration and segregation.

Conclusion, limitations and future. Here, we adopt a big data approach to systematically explore the reliability of functional brain networks by richly sampling the parameters of various steps in the network construction and analysis pipeline. The results of this analysis provided robust experimental evidence supporting four key principles that will support reliable network neuroscience measurements and applications. These principles can serve as the base for building guidelines on the use of network neuroscience with rfMRI to map individual differences. Standard guidelines are essential for improvements of reproducibility in the research practice, and our findings provide experimental resources for such standardization in future network neuroscience applications. We note, however, that while our approach was extensive, it was not exhaustive – the analytical sampling procedure could miss many other existing choices (e.g., consensus-based thresholding for the edge filtering stage). The processing decisions that yield reliable connectomes may yield the most reliable network statistics, but there may be another way to process data that yields overall a higher level of reliability in network measures. Future work can build on our study by exploring these and other choices using the online computation and evaluation platform that accompanies the present study. Of note, the measurement reliability is not the final goal but the measurement validity, which must be considered although not easily ready for a direct examination (15). Validation (through various indirect validity assessments) on the use of the proposed principles represents a promising arena for future studies in functional network neuroscience (4).

Materials and Methods

Using the HCP test-retest dataset, our analytic procedure implemented four post-processing stages (Fig. 1a): node definition, edge construction, network measurement and reliability assessments. Specifically, the test-retest rfMRI dataset underwent the standardized preprocessing pipeline developed by the HCP team (32). The second step defines nodes (green box) using sets of brain areas based on 30 partitions, and then extracts the nodal time series. During the third step (yellow box), individual correlation matrices are first estimated based upon the six frequency bands derived from Buzsáki's theoretical framework on brain oscillations (33) along with the

classical band widely used (0.01 - 0.08 Hz). These matrices are then converted into adjacency matrices using $4 \times 12 = 48$ strategies on edge filtering. In the fourth step, we performed graph analyses (blue box) by systematically calculating the brain graph metrics at global, modular and nodal scales. Finally, test-retest reliability was evaluated (red box) as ICCs with the linear mixed models.

Test-Retest Dataset. The WU-Minn Consortium in HCP shared a set of test-retest multimodal MRI datasets of 46 subjects from both the S1200 release and the Retest release. These subjects were retested using the full HCP 3T multimodal imaging and behavioral protocol. Each subject underwent the four scans on two days (two scans per day: REST1 versus REST2) during the first visit and returned several months later to finish the four scans on another two days during the second visit (Fig. 1b). The test-retest interval ranged from 18 to 328 days (mean: 4.74 months, standard deviation: 2.12 months). Only 41 subjects (28 females, age range: 26-35 years; 13 males, age range: 22-33 years) had full length rfMRI data across all the eight scans, and were included in the subsequent analyses. As indicated in the literature (21, 32), rfMRI protocols used by HCP for scanning and preprocessing images have been optimized for reliability.

During the scanning, participants were instructed to keep their eyes open and to let their mind wander while fixating on a cross-hair projected on a dark background. Data were collected on the 3T Siemens Connectome Skyra MRI scanner with a 32-channel head coil. All functional images were acquired using a multiband gradient-echo EPI imaging sequence [2mm isotropic voxel, 72 axial slices, TR = 720ms, TE = 33.1ms, flip angle = 52°, field of view = 208×180 mm^2, matrix size = 104×90 and a multiband factor of 8]. A total of 1200 images was acquired for a duration of 14 min and 24 s. Details on the imaging protocols can be found in (76).

The protocols of rfMRI image preprocessing and artifact-removal procedures are documented in detail elsewhere and generated the minimally preprocessed HCP rfMRI images. It is note that artifacts were removed using the Oxford Center for Functional MRI of the Brain's ICA-based X-noiseifier (ICA + FIX) procedure, followed by MS-MAll for inter-subject registration. The preprocessed rfMRI data were represented as a time series of grayordinates (4D), combining both cortical surface vertices and subcortical voxels (32).

Node Definition. A brain graph defines a node as a brain area, which is generally derived by an element of brain parcellation (parcel) according to borders or landmarks of brain anatomy, structure or function as well as an element of volume (voxel) in imaging signal acquisition or a cluster of voxels (77). Due to the high computational demand of voxel-based brain graph, in this study we defined nodes as parcels according to the following brain parcellation strategies (Fig. 2a). A surface-based approach has been demonstrated to outperform other approaches for functional MRI analysis (24, 78) and thus the nodes are defined in the surface space (total 30 surface parcellation choices). We adopted a naming convention for brain parcellations as follows: '*ParcAbbr-NumberOfParcels*' (e.g., LGP-100 or its whole-brain version wbLGP-458).

HCP Multi-Modal Parcellation (MMP) A cortical parcellation generated from multi-modal images of 210 adults from the HCP database, using a semi-automated approach (79). Cortical regions are delineated with respect to their function, connectivity, cortical architecture, and topography, as well as, expert knowledge and meta-analysis results from the literature (79). The HCP-MMP atlas contains 360 parcels (180 per hemisphere).

Local-Global Parcellation (LGP) A gradient-weighted Markov Random Field model integrating local gradient and global similarity approaches produces the novel parcellations (80). The final version of LGP comes with a multi-scale cortical atlas including 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000 parcels (equal numbers across the two hemispheres). One benefit of using LGP is to have nodes with almost the same size, and these nodes are also assigned to the common large-scale functional networks (38).

Brainnetome Parcellation (BNP) Both anatomical landmarks and connectivity-driven information are employed to develop this volumetric brain parcellation (81). Specifically, anatomical regions defined as in (82) are parcellated into subregions using functional and structural connectivity fingerprints from HCP datasets. Cortical parcels are obtained by projecting their volume space to surface space. It is noticed that the original BNP contains both cortical (105 areas per hemisphere) and subcortical (36 areas) regions but only the 210 cortical parcels are included for the subsequent analyses.

Whole-Brain Parcellation (wb) Inclusion of subcortical areas has been shown unignorable influences on brain graph analyses (22, 58), and we thus also constructed brain graphs with subcortical structures in volume space as nodes by adding these nodes to the cortical brain graphs. To get a high-resolution subcortical parcellation, we adopted the 358 subcortical parcels in (83). The authors employed data of 337 unrelated HCP healthy volunteers and extended the MMP cortical network partition into subcortex. This results a set of whole-brain parcellations by combining these subcortical parcels with the aforementioned cortical parcellations, namely **wbMMP**,**wbLGP** and **wbBNP**. We noticed that the wbMMP-718 has been named by the authors of (83) as the Cole-Anticevic Brain-wide Network Partition, and we thus renamed the wbMMP-718 as wbCABP-718 for consistency.

We finally consider a whole-brain 'parcellation' derived by the group-level spatial independent component analysis (ICA) using the MELODIC (84) in FSL (**wbGICAP**) from 820 HCP subjects. The spatial ICA was applied in grayordinate space (surface vertices

plus subcortical grey matter voxels) and ran with different dimensionalities (15, 25, 50, 100, 200, 300), generating the set of spatial maps, which are considered as 'nodes'.

Edge Construction. After defining the node with each parcellation, in each parcel, regional mean time series were estimated by averaging the vertex time series at each time point. To construct an edge between a pair of nodes, their representative time series entered into the following steps in order: *band-pass filtering*, *inter-node connectivity transformation*, and *edge filtering*.

Band-Pass Filtering. Resting-state functional connectivity studies have typically focused on fluctuations below 0.08 Hz or 0.1 Hz (5, 85), and assumed that only these frequencies contribute significantly to inter-regional functional connectivity (FC) while other frequencies are artifacts (86). In contrast, however, other studies have found that specific frequency bands of the rfMRI oscillations make unique and neurobiologically meaningful contributions to resting-state functional connectivity (21, 87). More recently, with fast fMRI methods, some meaningful FC patterns were reported across much higher frequency bands (88). These observations motivate exploring a range of frequency bands beyond those typically studied in resting-state functional connectivity studies, including faster frequencies.

Buzsáki and Draguhn (33) proposed a hierarchical organization of frequency bands driven by the natural logarithm linear law. This offers a theoretical template for partitioning rfMRI frequency content into multiple bands (Fig. 3a). The frequencies occupied by these bands have a relatively constant relationship to each other on a natural logarithmic scale and have a constant ratio between any given pair of neighboring frequencies (35). These different oscillations are linked to different neural activities, including cognition, emotion regulation, and memory (35, 62, 85). Advanced by the fast imaging protocols offered by the HCP scanner, the short scan interval (TR = 720ms) allows us to obtain more oscillation classes that the traditional rfMRI method. We incorporate the Buzsáki's framework (33, 36) with the HCP fast-TR datasets by using the DREAM toolbox (37) in the Connectome Computation System (27). It decomposed the time series into the six slow bands as illustrated in Fig. 3a.

Connectivity Transformation. For each scan, individual nodal representative time series were band-pass filtered with each of the six frequency bands, and another empirical frequency band, slow-emp (0.01-0.08Hz). The Pearson's correlation $r_{ij} \in [-1,1]$ between the filtered time series of each pair of nodes $i = 1,...,N, j = 1,...,N$ was calculated (N is the number of nodes). These correlation values provided an estimation on the edge strengths between the two nodes, and formed a $N \times N$ symmetric correlation matrix $R = (r_{ij})$ for each given subject, scan, parcellation, and frequency band.

Many network metrics are not well defined for negatively weighted connections. In order to ensure that the connection weights are positive only, we applied four types of transformations to the symmetric correlation matrix: the **positive** (Eq.pos), **absolute** (Eq.abs), **exponential** (Eq.exp) and **distance-inverse** (Eq.div) functions, respectively. This avoids the negative values in the inter-node connectivity matrix $W = (w_{ij})$ where $z_{ij} = \tanh^{-1}(r_{ij})$ is Fisher's $z-$transformation.

$$w_{ij} = \frac{z_{ij} + |z_{ij}|}{2} \in [0, \infty) \qquad \textbf{(pos)}$$

$$w_{ij} = |z_{ij}| \in [0, \infty) \qquad \textbf{(abs)}$$

$$w_{ij} = \mathbf{e}^{z_{ij}} \in [0, \infty) \qquad \textbf{(exp)}$$

$$w_{ij} = \frac{2}{\sqrt{2 \times (1 - r_{ij})}} \in (0, \infty) \qquad \textbf{(div)}$$

The connectivity matrix represents a set of the node parcels and relational quantities between each pair of the nodes, and will serve as the basis of following edge filtering procedure for generation of the final brain graphs.

Edge Filtering. In a graph, edges represent a set of relevant interactions of crucial importance to obtain parsimonious descriptions of complex networks. Filtering valid edges can be highly challenging due to the lack of 'ground truth' of the human brain connectome. To provide a reliable way of building candidate edges, we sampled the following 12 schemes on edge filtering and applied them to the connectivity matrices.

Absolute Weight Thresholding (ABS) This approach selects those edges that exceed a manually defined absolute threshold (e.g., correlations higher than 0.5), setting all correlations smaller than 0.5 to 0 (ABS_{05}). This is a simple approach to reconstruct networks (89).

Proportional Thresholding (PROP) It is a common step in the reconstruction of functional brain networks to ensure equal edge density across subjects (90–92). It keeps the number of connections fixed across all individuals to rule out the influence of network density on the computation and comparison of graph metrics across groups. This approach includes the selection of a fixed percentage of the strongest conncections as edges in each individual network or brain graph. Compared to ABS, PROP has been argued to reliably separate density from topological effects (28, 93) and to result in more stable network metrics (94). This makes it a commonly used approach for network construction and analysis in disease-related studies. Here, we focused on two threshholds that are commonly reported in the literature: 10% ($PROP_{10}$) and 20% ($PROP_{20}$).

Degree Thresholding (DEG) The structure of a graph can be biased by the number of existing edges. Accordingly, statistical measures derived from the graph should be compared against graphs that preserve the same average degree, K. A threshold of the degree

can be chosen to produce graphs with a fixed mean degree (e.g., $K = 5$, DEG_5), which is the average nodal degrees of an individual graph from a single subject's scan. Many network neuroscience studies have taken this choice for $K = 5$ (95–98). We also include the DEG_{15} for denser graphs of the brain networks.

Global Cost Efficiency Optimization (GCE) Given a network with a cost ρ, its global efficiency is a function of the cost $E_g(\rho)$, and its GCE is $J(\rho) = E_g(\rho) - \rho$. Several studies suggested that brain networks, in particular those with small-world topology, maximize their global-cost efficiency (43), i.e., $J^{max} = \max_\rho J(\rho)$. Computationally, this scheme is implemented by looping all network costs (e.g., adding edges with weights in order) to find the J^{max} (see Fig. 2b) where the corresponding edge weight was determined as the threshold for edge filtering. In this sense, GCE is an individualised and optimised version of ABS, PROP and DEG while the latter three are commonly employed with a fixed threshold for all individuals.

Overall Efficiency Cost Optimization (ECO) Both global and local efficiency are important graph features to characterize the structure of complex systems in terms of integration and segregation of information (99). ECO was proposed to determine a network density threshold for filtering out the weakest links (100). It maximizes an extension of J^{max}, the ratio between the overall (both global and local) efficiency and its wiring cost $\max_\rho J^{ext}(\rho) = (E_g(\rho) + E_{loc}(\rho))/\rho$ where E_{loc} denotes the network local efficiency. The study (99) also demonstrated that, to maximize J, these networks have to be sparse with an average node degree $K \simeq 3$.

Minimum Spanning Tree (MST) This is an increasingly popular method for identifying the smallest and most essential set of connections while ensuring that the network forms a fully connected graph (101–104). The tenet of using MST is to summarize information and index structure of the graph, and thus remove edges with redundant information (39). Specifically, an MST filtered graph will contain N nodes connected *via* $N-1$ connections with minimal cost and no loops. This addresses key issues in existing topology filtering schemes that rely on arbitrary and user-specified absolute thresholds or densities.

Orthogonal Minimum Spanning Tree (OMST) This topological filtering scheme was proposed recently (40) to maximize the information flow over the network *versus* the cost by selecting the connections via the OMSTs. It samples the full-weighted brain network over consecutive rounds of MST that are orthogonal to each other (see Fig. 2b). Practically, we extracted the 1st MST, and then we cleared their connections and we tracked the 2nd MST from the rest of the network connections, etc. Such an iterative procedure (stopped by the Mth MST) can get orthogonal MSTs and topologically filter brain network by optimizing the GCE under the constrains by the MST, leading to an integration of both GCE and MST

$$\max_{n\in[1,M]} J(\rho(n\text{MSTs})) = E_g(\rho(n\text{MSTs})) - \rho(n\text{MSTs})$$

Planar Maximally Filtered Graph (PMFG) The idea underneath PMFG (41) is to filter a dense matrix of weights by retaining the largest possible subgraph while imposing global constraints on the topology of the resulting network. In particular, edges with the strong connection weights are retained while constraining the subgraph to be a (spanning) tree globally. Similarly, during the PMFG construction, the largest weights are retained while constraining the subgraph to be a planar graph globally. The PMFG algorithm searches for the maximum weighted planar subgraph by adding edges one by one. The resulting matrix is sparse, with $3(N-2)$ edges. It starts by sorting all the edges of a dense matrix of weights in non-increasing order and tries to insert every edge in the PMFG. Edges that violate the planarity constraint are discarded.

Triangulated Maximally Filtered Graph (TMFG) The algorithm for implementing PMFG is computationally expensive, and is therefore impractical when applied to large brain networks (42). A more efficient algorithms, TMFG, was developed that exhibited greatly reduced computational complexity compared to PMFG. This method captures the most relevant information between nodes by approximating the network connectivity matrix with the endorsement association matrix and minimizing spurious associations. The TMFG derived network contains 3-node (triangle) and 4-node (tetrahedron) cliques, imposing a nested hierarchy and automatically generates a chordal network (42, 105). Although TMFG is not widely applied in network neuroscience studies, it as been applied elsewhere and proven to be a suitable choice for modeling interrelationships between psychological constructs like personality traits (106).

Orthogonal TMF Graph (OTMFG) To combine both the TMFG's efficiency and OMST's accuracy, we propose OTMFG to maximize the information flow over the network *versus* the cost by selecting the connections of the orthogonal TMFG. It samples the full-weighted brain network over consecutive rounds of TMFG that are orthogonal to each other.

In summary, as illustrated in Fig. 4a, the 12 edge filtering schemes transform a fully weighted matrix into a sparse matrix to represent the corresponding brain network. They can be categorized into two classes: threshold-based *versus* topology-based schemes. ABS_{05}, PRO_{10}, PRO_{20}, DEG_5, DEG_{15}, ECO and GCE rely on a threshold for filtering and retaining edges with higher weights than the threshold. These

schemes normally ignore the topological structure of the entire network and can result in isolated nodes. In contrast, the topology-based methods including MST, OMST, PMFG, TMFG and OTMFG, all consider the global network topology in determining which edges to retain. As illustrated in Fig. 4b, all the schemes are plotted in the $\rho - J^{max}$ plane for their network economics.

Network Analysis. We performed graph-theory-driven network analysis by calculating several common graph-based metrics for the resulting graphs. These measures, broadly, can be interpreted based on whether the characterize the extent to which network structure allows for integrated or segregation information flow. Examples of integrative measures include average shortest path length (L_p), global efficiency (E_g), and pseudo diameter (D). Segregative measures include clustering coefficient (C_p), local efficiency (E_{local}), transitivity (Tr), modularity (Q), and a suite of nodal centrality measures (Appendix 1). All the metrics are calculated using functions included in the Brain Connectivity Toolbox (107). To improve computation, we also used Graph-tool (108) and NetworKit (109) which achieve performance comparable (both in memory usage and computation time) to that of a pure C/C++ library. We treated these metrics as the network measurements for subsequent reliability analysis.

Reliability Assessments. Measurement reliability is defined as the extent to which measurements can be replicated across multiple repeated measures. Test-retest reliability is the closeness of the agreement between the results of successive measurements of the same measure and carried out under the same conditions of measurement.

Linear mixed models. As a group-level statistic, reliability refers to the inter-individual or between-subject variability V_b relative to the intra-individual or within-subject variability V_w. Both the intra- and inter-individual variances can be estimated using linear mixed model (LMM). In this study, given a functional graph metric ϕ, we considered a random sample of P subjects with N repeated measurements of a continuous variable in M visits. ϕ_{ijk} (for $i = 1, \cdots, N$ and $j = 1, \cdots, M$, and $k = 1, \cdots, P$) denotes the metric from the kth subject's j^{th} visit and i^{th} measurement occasions. The three-level LMM models ϕ_{ijk} as the following equations:

$$\overbrace{\phi_{ijk}}^{\text{Graph metric}} = \underbrace{\gamma_{000}}_{\substack{\text{fixed}\\\text{intercept}}} + \underbrace{p_{0k}}_{\substack{\text{random intercepts}\\\text{level 3, subjects}}} + \underbrace{v_{0jk}}_{\substack{\text{random intercepts}\\\text{level 2, visits}}} + \underbrace{e_{ijk}}_{\substack{\text{random}\\\text{residuals}}}$$

Where γ_{000} is a fixed parameter (the group mean) and p_{0k}, v_{0jk} and e_{ijk} are independent random effects normally distributed with a mean of 0 and variances σ^2_{p0}, σ^2_{v0}, and σ^2_e. The term p_{0k} is the subject effect, v_{0jk} is the visit effect and e_{ijk} is the measurement residual. Age, gender and interval(Δt) between two visits are covariants.

ICC Estimation. These variances are used to calculate the test-retest reliability, which is measured by the dependability coefficient and reflects the absolute agreement of measurements. The dependability coefficient is a form of the intraclass correlation coefficient (ICC) commonly, which is the ratio of the variances due to the object of measurement versus sources of error. To avoid negative ICC values and obtain more accurate estimation of the sample ICC, the variance components in model are usually estimated with the restricted maximum likelihood (ReML) approach with the covariance structure of an unrestricted symmetrical matrix (24).

$$\text{Reliability}(\phi) = \frac{V_b}{V_b + V_w} = \frac{\sigma^2_{p0}}{\sigma^2_{p0} + \sigma^2_e} \qquad \textbf{(ICC)}$$

A metric with moderate to almost perfect test-retest reliability (ICC ≥ 0.4) is commonly expected in practice. The level of reliability should not be judged only based upon the point statistical estimation of ICC but its confidence intervals (CI) (110). We employed the nonparametric conditional bootstrap method for 1000 times to estimate their 95% CIs.

Statistics Evaluation. Our analyses can produce big data of reliability statistics including 524,160 ICCs for the global network metrics. These ICCs are grouped into four categories (parcellation, frequency band, connectivity transformation and edge filtering scheme), each of which has different choices. Given each choice of a category, we estimated its density distributions of ICCs and calculated two descriptive statistics: 1) mean ICC values, which measures the *general reliability* under the given choice; 2) number of almost perfect (noap) ICC values, which measures the *potential reliability* under the given choice.

We further perform Friedman rank sum test to evaluate whether the location parameters of the distribution of ICCs are the same in each choice. Once the Friedman test is significant, we employ the pairwise Wilcoxon signed rank test for post-hoc evaluations to compare ICCs between each pair of the distributions under different choices. The statistical significance levels are corrected with Bonferroni method for controlling the family wise error rate at a level of 0.05. We develop a method to visualize and evaluate the change of ICCs (i.e., reliability gradient) between different choices (Fig. 1c). Specifically, the reliability can be plotted as a function of V_b and V_w in its anatomy plane (14, 15). The gradient of reliability between two choices is modeled by the vector (i.e., the black arrow), and decomposed into changes of individual variability. The systematic evaluation on the reliability of the global network metrics determines the optimal network neuroscience by combining the most reliable pipeline choices. Finally, the optimized pipeline generates the nodal metrics as well as their reliability.

ACKNOWLEDGEMENTS

This work was supported in part by the Startup Funds for Leading Talents at Beijing Normal University, the National Basic Science Data Center 'Chinese Data-sharing Warehouse for In-vivo Imaging Brain' (NBSDC-DB-15), Beijing Municipal Science and Technology Commission (Z161100002616023, Z181100001518003), the Key-Area Research and Development Program of Guangdong Province (2019B030335001), Guangxi BaGui Scholarship (201621), and Indiana University Office of the Vice President for Research Emerging Area of Research Initiative, Learning: Brains, Machines and Children. The neuroimaging data were provided by the HCP WU-Minn Consortium, which is funded by the 16 NIH institutes and centers that support the NIH Blueprint for Neuroscience Research 1U54MH091657 (PIs: David Van Essen and Kamil Ugurbil), the McDonnell Center for Systems Neuroscience at Washington University.

Bibliography

1. D.S. Bassett, K.E. Cullen, S.B. Eickhoff, M.J. Farah, Y. Goda, P. Haggard, H. Hu, Y.L. Hurd, S.A. Josselyn, B.S. Khakh, J.A. Knoblich, P. Poirazi, R.A. Poldrack, M. Prinz, P.R. Roelfsema, T.L. Spires-Jones, M. Sur, and H.R. Ueda. Reflections on the past two decades of neuroscience. *Nature Reviews Neuroscience*, 21(10):524–534, 2020.
2. E. Bullmore and O. Sporns. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.
3. D.S. Bassett and O. Sporns. Network neuroscience. *Nature Neuroscience*, 20(3):353–364, 2017.
4. D.S. Bassett, P. Zurn, and J.I. Gold. On the nature and use of models in network neuroscience. *Nature Reviews Neuroscience*, 19(9):566–578, 2018.
5. B. Biswal, F. Zerrin Yetkin, V.M. Haughton, and J.S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, 34(4):537–541, 1995.
6. B.B. Biswal, M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S.M. Smith, C.F. Beckmann, J.S. Adelstein, R.L. Buckner, S. Colcombe, A.-M. Dogonowski, M. Ernst, D. Fair, M. Hampson, M.J. Hoptman, J.S. Hyde, V.J. Kiviniemi, R. Kötter, S.-J. Li, C.-P. Lin, M.J. Lowe, C. Mackay, D.J. Madden, K.H. Madsen, D.S. Margulies, H.S. Mayberg, K. McMahon, C.S. Monk, S.H. Mostofsky, B.J. Nagel, J.J. Pekar, S.J. Peltier, S.E. Petersen, V. Riedl, S.A.R.B. Rombouts, B. Rypma, B.L. Schlaggar, S. Schmidt, R.D. Seidler, G.J. Siegle, C. Sorg, G.-J. Teng, J. Veijola, A. Villringer, M. Walter, L. Wang, X.-C. Weng, S. Whitfield-Gabrieli, P. Williamson, C. Windischberger, Y.-F. Zang, H.-Y. Zhang, F.X. Castellanos, and M.P. Milham. Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10):4734–4739, 2010.
7. C. Kelly, B.B. Biswal, R.C. Craddock, F.X. Castellanos, and M.P. Milham. Characterizing variation in the functional connectome: Promise and pitfalls. *Trends in Cognitive Sciences*, 16(3):181–188, 2012.
8. S.M. Smith, D. Vidaurre, C.F. Beckmann, M.F. Glasser, M. Jenkinson, K.L. Miller, T.E. Nichols, E.C. Robinson, G. Salimi-Khorshidi, M.W. Woolrich, D.M. Barch, K. Uğurbil, and D.C. Van Essen. Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*, 17(12):666–682, 2013.
9. K.R.A. Van Dijk, T. Hedden, A. Venkataraman, K.C. Evans, S.W. Lazar, and R.L. Buckner. Intrinsic functional connectivity as a tool for human connectomics: Theory, properties, and optimization. *Journal of Neurophysiology*, 103(1):297–321, 2010.
10. J. Dubois and R. Adolphs. Building a science of individual differences from fMRI. *Trends in Cognitive Sciences*, 20(6):425–443, 2016.
11. X.-N. Zuo, Y. He, R.F. Betzel, S. Colcombe, O. Sporns, and M.P. Milham. Human connectomics across the life span. *Trends in Cognitive Sciences*, 21(1):32–45, 2017.
12. L. Pessoa. Understanding emotion with brain networks. *Current Opinion in Behavioral Sciences*, 19:19–25, 2018.
13. A. Fornito, A. Zalesky, and M. Breakspear. The connectomics of brain disorders. *Nature Reviews Neuroscience*, 16(3):159–172, 2015.
14. X.-X. Xing and X.-N. Zuo. The anatomy of reliability: a must read for future human brain mapping. *Science Bulletin*, 63(24):1606–1607, 2018.
15. X.-N. Zuo, T. Xu, and M.P. Milham. Harnessing reliability for neuroscience research. *Nature Human Behaviour*, 3(8):768–771, 2019.
16. Zeyi Wang, Eric Bridgeford, Shangsi Wang, Joshua T. Vogelstein, and Brian Caffo. Statistical analysis of data repeatability measures, 2020.
17. M.L. Elliott, A.R. Knodt, A. Caspi, T.E. Moffitt, and A.R. Hariri. Need for psychometric theory in neuroscience research and training: Reply to kragel et al. (2021). *Psychological Science*, 2021.
18. H.C. Kraemer. The reliability of clinical diagnoses: State of the art. *Annual Review of Clinical Psychology*, 10:111–130, 2014.
19. M. Herting, P. Gautam, Z. Chen, A. Mezher, and N.C. Vetter. Test-retest reliability of longitudinal task-based fmri: Implications for developmental studies. *Developmental Cognitive Neuroscience*, 33:17–26, 2018.
20. P.M. Matthews and A. Hampshire. Clinical concepts emerging from fMRI functional connectomics. *Neuron*, 91(3):511–528, 2016.
21. X.-N. Zuo and X.-X. Xing. Test-retest reliabilities of resting-state fmri measurements in human brain functional connectomics: A systems neuroscience perspective. *Neuroscience and Biobehavioral Reviews*, 45:100–118, 2014.
22. S. Noble, D. Scheinost, and R.T. Constable. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*, 203, 2019.
23. Michael N. Hallquist and Frank G. Hillary. Graph theory approaches to functional network organization in brain disorders: A critique for a brave new small-world. *Network Neuroscience*, 3(1):1–26, 2019.
24. X.-N. Zuo, T. Xu, L. Jiang, Z. Yang, X.-Y. Cao, Y. He, Y.-F. Zang, F.X. Castellanos, and M.P. Milham. Toward reliable characterization of functional homogeneity in the human brain: Preprocessing, scan duration, imaging resolution and computational space. *NeuroImage*, 65:374–386, 2013.
25. J. Wang, L. Wang, Y. Zang, H. Yang, H. Tang, Q. Gong, Z. Chen, C. Zhu, and Y. He. Parcellation-dependent small-world brain functional networks: A resting-state fMRI study. *Human Brain Mapping*, 30(5):1511–1523, 2009.
26. X. Liang, J. Wang, C. Yan, N. Shu, K. Xu, G. Gong, and Y. He. Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: A resting-state functional mri study. *PLoS One*, 7(3), 2012.
27. T. Xu, Z. Yang, L. Jiang, X.-X. Xing, and X.-N. Zuo. A Connectome Computation System for discovery science of brain. *Science Bulletin*, 60(1):86–95, 2015.
28. U. Braun, M.M. Plichta, C. Esslinger, C. Sauer, L. Haddad, O. Grimm, D. Mier, S. Mohnke, A. Heinz, S. Erk, H. Walter, N. Seiferth, P. Kirsch, and A. Meyer-Lindenberg. Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *NeuroImage*, 59(2):1404–1412, 2012.
29. M. Termenon, A. Jaillard, C. Delon-Martin, and S. Achard. Reliability of graph analysis of resting state fMRI using test-retest dataset from the Human Connectome Project. *NeuroImage*, 142:172–187, 2016.
30. N.K. Aurich, J.O.A. Filho, A.M.M. da Silva, and A.R. Franco. Evaluating the reliability of different preprocessing steps to estimate graph theoretical measures in resting state fMRI data. *Frontiers in Neuroscience*, 9(FEB), 2015.
31. D.C. Van Essen, S.M. Smith, D.M. Barch, T.E.J. Behrens, E. Yacoub, and K. Ugurbil. The WU-Minn Human Connectome Project: An overview. *NeuroImage*, 80:62–79, 2013.
32. M.F. Glasser, S.N. Sotiropoulos, J.A. Wilson, T.S. Coalson, B. Fischl, J.L. Andersson, J. Xu, S. Jbabdi, M. Webster, J.R. Polimeni, D.C. Van Essen, and M. Jenkinson. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80:105–124, 2013.
33. G. Buzsaki and A. Draguhn. Neuronal oscillations in cortical networks. *Science*, 304 (5679):1926–1929, 2004. ISSN 0036-8075.
34. J.R. Landis and G.G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.
35. G. Buzsáki. *Rhythms of the Brain*. Oxford University Press, 2009.
36. M. Penttonen and G. Buzsáki. Natural logarithmic relationship between brain oscillators. *Thalamus and Related Systems*, 2(2):145–152, 2003.
37. Z.-Q. Gong, P. Gao, C. Jiang, X.-X. Xing, H.-M. Dong, T. White, F.X. Castellanos, H.-F. Li, and X.-N. Zuo. Dream: A toolbox to decode rhythms of the brain system. *Neuroinformatics*, 2021.
38. B.T. Thomas Yeo, F.M. Krienen, J. Sepulcre, M.R. Sabuncu, D. Lashkari, M. Hollinshead, J.L. Roffman, J.W. Smoller, L. Zöllei, J.R. Polimeni, B. Fisch, H. Liu, and R.L. Buckner. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125–1165, 2011.
39. R.N. Mantegna. Hierarchical structure in financial markets. *European Physical Journal B*, 11(1):193–197, 1999.
40. S.I. Dimitriadis, M. Antonakakis, P. Simos, J.M. Fletcher, and A.C. Papanicolaou. Data-driven topological filtering based on orthogonal minimal spanning trees: Application to multigroup magnetoencephalography resting-state connectivity. *Brain Connectivity*, 7(10): 661–670, 2017.
41. M. Tumminello, T. Aste, T. Di Matteo, and R.N. Mantegna. A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10421–10426, 2005.
42. G.P. Massara, T. Di Matteo, and T. Aste. Network filtering for big data: Triangulated maximally filtered graph. *Journal of Complex Networks*, 5(2):161–178, 2016.
43. D S Bassett, E Bullmore, B A Verchinski, V S Mattay, D R Weinberger, and A Meyer-Lindenberg. Hierarchical Organization of Human Cortical Networks in Health and Schizophrenia. *Journal of Neuroscience*, 28(37):9239–9248, 2008.
44. S.B. Eickhoff, B. Thirion, G. Varoquaux, and D. Bzdok. Connectivity-based parcellation: Critique and implications. *Human Brain Mapping*, 36(12):4771–4792, 2015.
45. A. Fornito, A. Zalesky, and M. Breakspear. Graph analysis of the human connectome: Promise, progress, and pitfalls. *NeuroImage*, 80:426–444, 2013.
46. A. Zalesky, A. Fornito, I.H. Harding, L. Cocchi, M. Yücel, C. Pantelis, and E.T. Bullmore. Whole-brain anatomical networks: Does the choice of nodes matter? *NeuroImage*, 50(3): 970–983, 2010.
47. X.-N. Zuo, B.B. Biswal, and R.A. Poldrack. Editorial: Reliability and reproducibility in functional connectomics. *Frontiers in Neuroscience*, 13(FEB), 2019.
48. G.E. Alexander, M.R. DeLong, and P.L. Strick. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, 9:357–381, 1986.
49. P.L. Strick. How do the basal ganglia and cerebellum gain access to the cortical motor areas? *Behavioural Brain Research*, 18(2):107–123, 1985.
50. A.C. Bostan and P.L. Strick. The basal ganglia and the cerebellum: Nodes in an integrated network. *Nature Reviews Neuroscience*, 19(6):338–350, 2018.
51. M.G. Packard and B.J. Knowlton. Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience*, 25:563–593, 2002.
52. A. Di Martino, A. Scheres, D.S. Margulies, A.M.C. Kelly, L.Q. Uddin, Z. Shehzad, B. Biswal, J.R. Walters, F.X. Castellanos, and M.P. Milham. Functional connectivity of human striatum: A resting state fmri study. *Cerebral Cortex*, 18(12):2735–2747, 2008.
53. E.Y. Choi, B.T. Thomas Yeo, and R.L. Buckner. The organization of the human striatum estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 108(8):2242–2263, 2012.
54. R.L. Buckner, F.M. Krienen, A. Castellanos, J.C. Diaz, and B.T. Thomas Yeo. The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(5):2322–2345, 2011.
55. S. Marek, J.S. Siegel, E.M. Gordon, R.V. Raut, C. Gratton, D.J. Newbold, M. Ortega, T.O. Laumann, B. Adeyemo, D.B. Miller, A. Zheng, K.C. Lopez, J.J. Berg, R.S. Coalson, A.L. Nguyen, D. Dierker, A.N. Van, C.R. Hoyt, K.B. McDermott, S.A. Norris, J.S. Shimony, A.Z. Snyder, S.M. Nelson, D.M. Barch, B.L. Schlaggar, M.E. Raichle, S.E. Petersen, D.J. Greene, and N.U.F. Dosenbach. Spatial and temporal organization of the individual human cerebellum. *Neuron*, 100(4):977–993.e7, 2018.
56. K.A. Barnes, A.L. Cohen, J.D. Power, S.M. Nelson, Y.B.L. Dosenbach, F.M. Miezin, S.E.

Petersen, and B.L. Schlaggar. Identifying basal ganglia divisions in individuals using resting-state functional connectivity mri. *Frontiers in Systems Neuroscience*, 4, 2010.
57. M. Garcia-Garcia, A. Nikolaidis, P. Bellec, R.C. Craddock, B. Cheung, F.X. Castellanos, and M.P. Milham. Detecting stable individual differences in the functional organization of the human basal ganglia. *NeuroImage*, 170:68–82, 2018.
58. D.J. Greene, S. Marek, E.M. Gordon, J.S. Siegel, C. Gratton, T.O. Laumann, A.W. Gilmore, J.J. Berg, A.L. Nguyen, D. Dierker, A.N. Van, M. Ortega, D.J. Newbold, J.M. Hampton, A.N. Nielsen, K.B. McDermott, J.L. Roland, S.A. Norris, S.M. Nelson, A.Z. Snyder, B.L. Schlaggar, S.E. Petersen, and N.U.F. Dosenbach. Integrative and network-specific connectivity of the basal ganglia and thalamus defined in individuals. *Neuron*, 105(4):742–758.e6, 2020.
59. R. Sala-Llonch, S.M. Smith, M. Woolrich, and E.P. Duff. Spatial parcellations, spectral filtering, and connectivity measures in fMRI: Optimizing for discrimination. *Human Brain Mapping*, 40(2):407–419, 2019.
60. Usama Pervaiz, Diego Vidaurre, Mark W. Woolrich, and Stephen M. Smith. Optimising network modelling methods for fMRI. *NeuroImage*, 211:116604, 2020.
61. X.-N. Zuo, C. Kelly, J.S. Adelstein, D.F. Klein, F.X. Castellanos, and M.P. Milham. Reliable intrinsic connectivity networks: Test-retest evaluation using ICA and dual regression approach. *NeuroImage*, 49(3):2163–2177, 2010.
62. S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1):63–72, 2006.
63. Y. Wang, Q. Zou, Y. Ao, Y. Liu, Y. Ouyang, X. Wang, B. Biswal, Q. Cui, and H. Chen. Frequency-dependent circuits anchored in the dorsal and ventral left anterior insula. *Scientific Reports*, 10(1), 2020.
64. E. Bullmore and O. Sporns. The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336–349, 2012.
65. D.S. Bassett, B.G. Nelson, B.A. Mueller, J. Camchong, and K.O. Lim. Altered resting state complexity in schizophrenia. *NeuroImage*, 59(3):2196–2207, 2012.
66. M.W. Cole, T. Yarkoni, G. Repovš, A. Anticevic, and T.S. Braver. Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *Journal of Neuroscience*, 32 (26):8988–8999, 2012.
67. E. Santarnecchi, G. Galli, N.R. Polizzotto, A. Rossi, and S. Rossi. Efficiency of weak brain connections support general cognitive functioning. *Human Brain Mapping*, 35(9): 4566–4582, 2014.
68. J. Faskowitz, F.Z. Esfahlani, Y. Jo, O. Sporns, and R.F. Betzel. Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture. *Nature Neuroscience*, 2020.
69. T.S. Jackson and N. Read. Theory of minimum spanning trees. i. mean-field theory and strongly disordered spin-glass model. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 81(2), 2010.
70. X.-N. Zuo, R. Ehmke, M. Mennes, D. Imperati, F.X. Castellanos, O. Sporns, and M.P. Milham. Network centrality in the human functional connectome. *Cerebral Cortex*, 22(8): 1862–1875, 2012.
71. Mehraveh Salehi, Amin Karbasi, Daniel S Barron, Dustin Scheinost, and R Todd Constable. Individualized functional networks reconfigure with cognitive state. *NeuroImage*, 206: 116233, 2020.
72. P. Rakic. Evolution of the neocortex: A perspective from developmental biology. *Nature Reviews Neuroscience*, 10(10):724–735, 2009.
73. C.-H. Chen, E.D. Gutierrez, W. Thompson, M.S. Panizzon, T.L. Jernigan, L.T. Eyler, C. Fennema-Notestine, A.J. Jak, M.C. Neale, C.E. Franz, M.J. Lyons, M.D. Grant, B. Fischl, L.J. Seidman, M.T. Tsuang, W.S. Kremen, and A.M. Dale. Hierarchical genetic organization of human cortical surface area. *Science*, 335(6076):1634–1636, 2012.
74. C.-H. Chen, M. Fiecas, E.D. Gutiérrez, M.S. Panizzon, L.T. Eyler, E. Vuoksimaa, W.K. Thompson, C. Fennema-Notestine, D.J. Hagler Jr., T.L. Jernigan, M.C. Neale, C.E. Franz, M.J. Lyons, B. Fischl, M.T. Tsuang, A.M. Dale, and W.S. Kremen. Genetic topography of brain morphology. *Proceedings of the National Academy of Sciences of the United States of America*, 110(42):17089–17094, 2013.
75. J.A. Miller, S.-L. Ding, S.M. Sunkin, K.A. Smith, L. Ng, A. Szafer, A. Ebbert, Z.L. Riley, J.J. Royall, K. Aiona, J.M. Arnold, C. Bennet, D. Bertagnolli, K. Brouner, S. Butler, S. Caldejon, A. Carey, C. Cuhaciyan, R.A. Dalley, N. Dee, T.A. Dolbeare, B.A.C. Facer, D. Feng, T.P. Fliss, G. Gee, J. Goldy, L. Gourley, B.W. Gregor, G. Gu, R.E. Howard, J.M. Jochim, C.L. Kuan, C. Lau, C.-K. Lee, F. Lee, T.A. Lemon, P. Lesnar, B. McMurray, N. Mastan, N. Mosqueda, T. Naluai-Cecchini, N.-K. Ngo, J. Nyhus, A. Oldre, E. Olson, J. Parente, P.D. Parker, S.E. Parry, A. Stevens, M. Pletikos, M. Reding, K. Roll, D. Sandman, M. Sarreal, S. Shapouri, N.V. Shapovalova, E.H. Shen, N. Sjoquist, C.R. Slaughterbeck, M. Smith, A.J. Sodt, D. Williams, L. Zöllei, B. Fischl, M.B. Gerstein, D.H. Geschwind, I.A. Glass, M.J. Hawrylycz, R.F. Hevner, H. Huang, A.R. Jones, J.A. Knowles, P. Levitt, J.W. Phillips, N. Šestan, P. Wohnoutka, C. Dang, A. Bernard, J.G. Hohmann, and E.S. Lein. Transcriptional landscape of the prenatal human brain. *Nature*, 508(7495):199–206, 2014.
76. S.M. Smith, C.F. Beckmann, J. Andersson, E.J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D.A. Feinberg, L. Griffanti, M.P. Harms, M. Kelly, T. Laumann, K.L. Miller, S. Moeller, S. Petersen, J. Power, G. Salimi-Khorshidi, A.Z. Snyder, A.T. Vu, M.W. Woolrich, J. Xu, E. Yacoub, K. Uğurbil, D.C. Van Essen, and M.F. Glasser. Resting-state fMRI in the Human Connectome Project. *NeuroImage*, 80:144–168, 2013.
77. O. Sporns. Network attributes for segregation and integration in the human brain. *Current Opinion in Neurobiology*, 23(2):162–171, 2013.
78. T.S. Coalson, D.C. Van Essen, and M.F. Glasser. The impact of traditional neuroimaging methods on the spatial localization of cortical areas. *Proceedings of the National Academy of Sciences of the United States of America*, 115(27):E6356–E6365, 2018.
79. M.F. Glasser, T.S. Coalson, E.C. Robinson, C.D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C.F. Beckmann, M. Jenkinson, S.M. Smith, and D.C. Van Essen. A multimodal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016.
80. A. Schaefer, R. Kong, E.M. Gordon, T.O. Laumann, X.-N. Zuo, A.J. Holmes, S.B. Eickhoff, and B.T.T. Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, 28(9):3095–3114, 2018.
81. L. Fan, H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A.R. Laird, P.T. Fox, S.B. Eickhoff, C. Yu, and T. Jiang. The Human Brainnetome Atlas: A new brain atlas based on connectional architecture. *Cerebral Cortex*, 26(8):3508–3526, 2016.
82. R.S. Desikan, F. Ségonne, B. Fischl, B.T. Quinn, B.C. Dickerson, D. Blacker, R.L. Buckner, A.M. Dale, R.P. Maguire, B.T. Hyman, M.S. Albert, and R.J. Killiany. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, 2006.
83. J.L. Ji, M. Spronk, K. Kulkarni, G. Repovš, A. Anticevic, and M.W. Cole. Mapping the human brain's cortical-subcortical functional network organization. *NeuroImage*, 185:35–57, 2019.
84. C.F. Beckmann, M. DeLuca, J.T. Devlin, and S.M. Smith. Investigations into resting-state connectivity using independent component analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):1001–1013, 2005.
85. M.D. Fox and M.E. Raichle. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, 8(9):700–711, 2007.
86. D. Cordes, V.M. Haughton, K. Arfanakis, J.D. Carew, P.A. Turski, C.H. Moritz, M.A. Quigley, and M.E. Meyerand. Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *American Journal of Neuroradiology*, 22(7):1326–1333, 2001.
87. R. Salvador, J. Suckling, M.R. Coleman, J.D. Pickard, D. Menon, and E. Bullmore. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral Cortex*, 15(9):1332–2342, 2005.
88. R.N. Boubela, K. Kalcher, W. Huf, C. Kronnerwetter, P. Filzmoser, and E. Moser. Beyond noise: Using temporal ICA to extract meaningful information from high-frequency fMRI signal fluctuations during rest. *Frontiers in Human Neuroscience*, 7:168, 2013.
89. P. Hagmann, M. Kurant, X. Gigandet, P. Thiran, V.J. Wedeen, R. Meuli, and J.-P. Thiran. Mapping human whole-brain structural networks with diffusion mri. *PLoS ONE*, 2(7), 2007.
90. Martijn P van den Heuvel, Siemon C de Lange, Andrew Zalesky, Caio Seguin, BT Thomas Yeo, and Ruben Schmidt. Proportional thresholding in resting-state fmri functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. *NeuroImage*, 152:437–449, 2017.
91. DS Bassett, ET Bullmore, A Meyer-Lindenberg, JA Apud, DR Weinberger, and R Coppola. Cognitive fitness of cost-efficient brain functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, 106(28):11747–11752, 2009.
92. M. Rubinov, O. Sporns, C. van Leeuwen, and M. Breakspear. Symbiotic relationship between brain structure and dynamics. *BMC Neuroscience*, 10, 2009. doi: 10.1186/1471-2202-10-55.
93. Cedric E Ginestet, Thomas E Nichols, Ed T Bullmore, and Andrew Simmons. Brain network analysis: separating cost from topology using cost-integration. *PLoS One*, 6(7): e21570, 2011.
94. K A Garrison, D Scheinost, E S Finn, X Shen, and R T Constable. The (in)stability of functional brain network measures across thresholds. *NeuroImage*, 118:651–661, September 2015.
95. R Milo, S Shen-Orr, S Itzkovitz, N Kashtan, D Chklovskii, and U Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
96. C Stam, B Jones, G Nolte, M Breakspear, and P Scheltens. Small-World Networks and Functional Connectivity in Alzheimer's Disease. *Cerebral Cortex*, 17(1):92–99, 2006.
97. Stavros I Dimitriadis, Nikolaos A Laskaris, Yolanda Del Rio-Portilla, and George Ch Koudounis. Characterizing Dynamic Functional Connectivity Across Sleep Stages from EEG. *Brain Topography*, 22(2):119–133, 2009.
98. S Micheloyannis, E Pachou, C J Stam, M Breakspear, P Bitsios, M Vourkas, S Erimaki, and M Zervakis. Small-world networks and disturbed functional connectivity in schizophrenia. *Schizophr Res*, 87(1-3):60–66, 2006.
99. Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical Review Letters*, 87(19):198701, 2001.
100. F. De Vico Fallani, V. Latora, and M. Chavez. A topological criterion for filtering information in complex brain networks. *PLoS Computational Biology*, 13(1), 2017. doi: 10.1371/journal.pcbi.1005305.
101. J. Meier, P. Tewarie, and P. Van Mieghem. The union of shortest path trees of functional brain networks. *Brain Connectivity*, 5(9):575–581, 2015.
102. D. van Nieuwenhuizen, L. Douw, M. Klein, S.M. Peerdeman, J.J. Heimans, J.C. Reijneveld, C.J. Stam, and A. Hillebrand. Cognitive functioning and functional brain networks in postoperative who grade i meningioma patients. *Journal of Neuro-Oncology*, 140(3):605–613, 2018.
103. W.M. Otte, E. van Diessen, S. Paul, R. Ramaswamy, V.P. Subramanyam Rallabandi, C.J. Stam, and P.K. Roy. Aging alterations in whole-brain networks during adulthood mapped with the minimum spanning tree indices: The interplay of density, connectivity cost and life-time trajectory. *NeuroImage*, 109:171–189, 2015.
104. H. Guo, M. Qin, J. Chen, Y. Xu, and J. Xiang. Machine-learning classifier for patients with major depressive disorder: Multifeature approach based on a high-order minimum spanning tree functional brain network. *Computational and Mathematical Methods in Medicine*, 2017, 2017.
105. W.-M. Song, T. Di Matteo, and T. Aste. Hierarchical information clustering by means of topologically embedded graphs. *PLoS One*, 7(3), 2012.
106. A.P. Christensen, Y.N. Kenett, T. Aste, P.J. Silvia, and T.R. Kwapil. Network structure of the wisconsin schizotypy scales–short forms: Examining psychometric network filtering approaches. *Behavior Research Methods*, 50(6):2531–2550, 2018.
107. M. Rubinov and O. Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010.
108. Tiago P. Peixoto. The graph-tool python library. *figshare*, 2014. doi: 10.6084/m9.figshare.1164194.
109. Christian L Staudt, Aleksejs Sazonovs, and Henning Meyerhenke. Networkit: A tool suite for large-scale complex network analysis. *Network Science*, 4(4):508–530, 2016.
110. T.K. Koo and M.Y. Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2):155–163, 2016.

Supplementary Information B

Table 1. Brief descriptions of the network metrics examined

Level	Measure	Attribute	Character	Ref
Global	basic	degree	k	
		number of edge	n_e	
		density	d	
	integration	global efficiency	Eg	
		avg shortest path length	Lp_a	
		avg nodal lp sw	Lp_b	
		pseudo diameter	D	
	segregation	clusteringcoef btc	Cp_a	
		clusteringcoef gtool	Cp_b	
		local efficiency 1	$Eloc_1$	
		local efficiency 2	$Eloc_2$	
		modularity	Q	
		transitivity bct	Tr_a	
		transitivity gtool	Tr_b	
	centrality	avg betweenness	Bc	
		avg eigenvector	Ec	
		avg pagerank	Pc	
		avg subgraph	Sc	
	resilience	assortativity	r	
		scalar assortativity	r_s	
		synchronizability	S	
		avg resolvent	Rv	
Nodal	integration	nodal path length	Lp_a	
		local path length	Lp_b	
	segregation	clustering coeff bct	Cp_a	
		clustering coeff gtool	Cp_b	
		local efficiency 1	$Eloc_1$	
		local efficiency 2	$Eloc_2$	
		nodal efficiency	E_{nodal}	
	centrality	degree centrality	Dc	
		betweeness centrality	Bc	
		eigenvector centrality	Ec	
		pagerank centrality	Pc	
		subgraph centrality	Sc	
		resolvent centrality	Rc	