Functional connectomics from a “big data” perspective

Mingrui Xia\textsuperscript{a,b,c}, Yong He\textsuperscript{a,b,c,∗}

\textsuperscript{a} National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China
\textsuperscript{b} IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China
\textsuperscript{c} Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing 100875, China

\begin{abstract}
In the last decade, explosive growth regarding functional connectome studies has been observed. Accumulating knowledge has significantly contributed to our understanding of the brain’s functional network architectures in health and disease. With the development of innovative neuroimaging techniques, the establishment of large brain datasets and the increasing accumulation of published findings, functional connectomic research has begun to move into the era of “big data”, which generates unprecedented opportunities for discovery in brain science and simultaneously encounters various challenging issues, such as data acquisition, management and analyses. Big data on the functional connectome exhibits several critical features: high spatial and/or temporal precision, large sample sizes, long-term recording of brain activity, multidimensional biological variables (e.g., imaging, genetic, demographic, cognitive and clinic) and/or vast quantities of existing findings. We review studies regarding functional connectomics from a big data perspective, with a focus on recent methodological advances in state-of-the-art image acquisition (e.g., multiband imaging), analysis approaches and statistical strategies (e.g., graph theoretical analysis, dynamic network analysis, independent component analysis, multivariate pattern analysis and machine learning), as well as reliability and reproducibility validations. We highlight the novel findings in the application of functional connectomic big data to the exploration of the biological mechanisms of cognitive functions, normal development and aging and of neurological and psychiatric disorders. We advocate the urgent need to expand efforts directed at the methodological challenges and discuss the direction of applications in this field.
\end{abstract}

Background

The human brain contains billions of neurons; each neuron connects to thousands of neurons via synapses. Thus, the human brain is naturally organized into a complex network that enables highly efficient information communication. The biophysical signals of the brain (e.g., the neural electrical and blood oxygen-level dependent signals) can be collected by neuroimaging techniques, such as functional magnetic resonance imaging (fMRI) and electroencephalography/magnetoencephalography (EEG/MEG). Functional brain networks can be constructed by estimating the relations among regional signals at macro- or microscales; these relations are referred to as the functional connectome (Biswal et al., 2010; Kelly et al., 2012; Sporns et al., 2005). Recent studies have demonstrated non-trivial organization principles of the human functional connectome, including segregated and integrated organization and modularized functional systems, which are significantly associated with individual cognitive performance and dynamically reorganize throughout a life span and in disease situations (Bullmore and Sporns, 2009, 2012; Cao et al., 2015; He and Evans, 2010; Misic and Sporns, 2016; Petersen and Sporns, 2015; Stam, 2014). These findings expand our understanding of the functional architecture of the human brain in health and disease.

In the past decade, the explosive growth in this field has resulted in the accumulation of extremely large amounts of data in various forms, including multimodal neuroimaging scans, vast quantities of cognitive data, genetic and environmental variables, and clinical measurements. In response to the broad and brilliant prospects of brain science in terms of applications in medical and computer sciences, several large projects (e.g., the Human Connectome Project, HCP, and Enhancing Neuro Imaging Genetics through Meta-Analysis, ENIGMA, Table 1) that target the development and application of innovative technologies, which can delineate the functions of the brain and inspire novel approaches for the diagnosis and treatment of neuropsychiatric disorders, have been implemented. These projects comprise a new generation of massive and precise data, which will guide research on the functional connectome into a “big data” era.

\begin{thebibliography}

\end{thebibliography}

Correspondence to: National Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing 100875, China.

E-mail address: yong.he@bnu.edu.cn (Y. He).

http://dx.doi.org/10.1016/j.neuroimage.2017.02.031
Received 8 October 2016; Accepted 13 February 2017
Available online 14 February 2017
1053-8119/ © 2017 Elsevier Inc. All rights reserved.
Connectomic big data represents the highly precise brain connectivity networks that are derived from large-sample, long-term recording or high-resolution data or are constructed by digging for undiscovered patterns in the vast quantity of existing findings. These datasets are usually accompanied by individual genetic and phenotypic information. Connectomic big data has at least one of the following typical features: i) high spatial and/or temporal precision provided by advanced neuroimaging and neurophysiological techniques; ii) a large sample size collected across projects and/or centers; iii) a long-term continuous brain recording across time; iv) multiple dimensions of biological variables for individuals (e.g., imaging, physiological, genetic, cognitive and clinical); and v) numerous published reports that are suitable for meta-analysis. Advanced analytical and statistical strategies, including graph theoretical approaches, dynamic connectivity analysis, independent component analysis, multivariate pattern analyses and machine learning techniques are being applied to functional connectomic big data. A variety of research on this topic has provided new insight into the functional architecture of the human brain (Fig. 1).

In this paper, we will provide a comprehensive review regarding functional connectomics studies from a big data perspective. We
primarily focus on the recent advances in state-of-the-art data acquisition and analysis strategies and reliability and reproducibility validation approaches for functional connectomic big data. We discuss the emerging challenges and issues that urgently need to be resolved in this field.

The acquisition of functional connectomic big data

Currently, functional connectomic big data on the human brain can be obtained by the following innovative acquisition techniques and approaches.

High-spatial-resolution techniques

The majority of studies on functional connectomics have constructed the brain's functional networks with regionally defined nodes (network size: ~20–1000), which has the advantage of saving computational power (e.g., Achard et al., 2006; de Reus and van den Heuvel, 2013; He et al., 2009; Salvador et al., 2005). Recently, high-spatial resolution-functional brain networks that were constructed at the voxel level (network size: ~20,000) have been able to unveil detailed connectivity information, especially for heterogeneous regions, without prior bias for choices of different regional parcellations (e.g., Buckner et al., 2009; Du et al., 2015; Hayasaka and Laurienti, 2010; Liao et al., 2013; Zuo et al., 2012). Recently developed parallel fMRI sequences, such as multiband multislice echo-planar imaging (Feinberg et al., 2010; Moeller et al., 2010; Setsompop et al., 2012), have been applied to accelerate the imaging of brain activity with a high-spatial resolution. The spatial resolution of raw (images constructed from k-space) fMRI data can increase from 3~5 mm to less than 1 mm (network size: ~1,000,000) and provide more accurate brain activity signals that reflect the heterogeneity of brain regions. Advances in ultrahigh magnetic field strengths, such as 7 T, enable functional imaging with high signal-to-noise ratios (Ugurbil et al., 2003; Vaughan et al., 2001) and fewer partial volume effects compared with traditional field strengths (Vu et al., 2016). The spatial resolution of high-magnetic-field imaging can attain the columnar level, which is crucial for distinguishing functional activity from distinct subunits, such as cortical layers and columns (De Martino et al., 2015). These high-spatial resolutions in brain images would undoubtedly generate functional brain networks at fine scales, which produces functional connectomic big data. Fig. 2 illustrates the correspondence between nodal sizes and network sizes.

High-temporal-resolution techniques

Conventional BOLD fMRI data acquisition usually has a temporal

---

Additional notes:

- **Fig. 1.** Framework that illustrates functional connectomic big data. A) Functional connectomic big data can be obtained by several methods, such as collecting large-sample, high-resolution neuroimaging data or searching for undiscovered patterns in the vast quantity of existing findings. B) Several currently available approaches, such as graph theoretical analysis, dynamical network analysis, independent component analysis, machine learning and novel statistical strategies, have been adopted in the field of functional connectomic big data. C) Functional connectomic big data and relevant analytical and statistical approaches have been employed to explore functional network architectures that underlie human cognition and behaviors, the trajectory of the brain changes during a lifespan, and the biological mechanisms of the human brain in health and disease.

- **Fig. 2.** Relationship between imaging resolution and high-resolution network size. In a voxelwise network, as the imaging resolution increases from 30 to 0.5 mm, A) the number of network nodes increases from $10^2$ to $10^7$, and B) the number of network edges increases from $10^3$ to $10^{13}$.
resolution of ~2 s. Significant improvements in fMRI acquisition elicited by parallel imaging techniques increased the temporal resolution of the sampling rate from ~2 s to a subsecond scale (e.g., ~500 ms) (Feinberg et al., 2010; Moeller et al., 2010; Setsompop et al., 2012), which enables researchers to collect temporal information about brain activity and simultaneously capture broader frequencies for some sources of biophysical noise, such as respiration and cardiac oscillations (Birn et al., 2006; Liu, 2013; Murphy et al., 2013). Thus, the effect of these noises can be effectively reduced to produce purer signals of brain activity. The increased temporal resolution of fMRI imaging provides unique opportunities to probe inter-regional functional connectivity at relatively high frequencies (e.g., 0.2–0.3 Hz) (Liao et al., 2013). Note that these fast-sampled functional imaging data enable the construction of dynamic network big data to ascertain time-varying connectivity architectures during a very short period (Hutchison et al., 2013a; Liao et al., 2015; Zalesky et al., 2014), which are associated with different brain activity states (Allen et al., 2014; Calhoun et al., 2014).

However, BOLD fMRI has a low sampling rate in contrast to the neuronal processing of the brain. Electrophysiological and optical imaging techniques (e.g., EEG/MEG/ECoG/INIRS) enable researchers to capture brain activity signals at a very high temporal resolution (maximum of ~1000 Hz) with relative convenience. Thus, these techniques naturally have advantages in studying the dynamic functional connectome at different time scales (from milliseconds to days) (Chu et al., 2012; Kramer et al., 2011). The combination of fMRI with simultaneous electrophysiological recordings has significant potential for exploring the temporal and spatial characteristics of the functional connectome.

**Large sample datasets**

Another important resource for functional connectomic big data is the increasing number of imaging datasets that are publicly available. Several big projects with large samples (Table 1) have been performed or are currently being performed, which increases the number of brain imaging databases with different research purposes. There projects include the 1000 Functional Connectomes Project (Biswal et al., 2010), the Human Connectome Project (Van Essen et al., 2013), the Consortium for Reliability and Reproducibility (CORR) (Zuo et al., 2014), the Enhancing Neuro Imaging Genetics through Meta Analysis (ENIGMA) (Thompson et al., 2014), the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (Mueller et al., 2005), the Attention Deficit Hyperactivity Disorder 200 (ADHD-200) (Consortium, 2012), the National Database for Autism Research (NDAR), the UK Biobank (Miller et al., 2016) and the OpenfMRI (Poldrack et al., 2013). Not including a large sample, the MyConnectome Project (Poldrack et al., 2015) gathered multidimensional information from one single subject over a period of 18 months (including multimodal imaging data, neuropsychological function, physical health, gene expression and metabolomics, as shown in Table 1), which provides a unique opportunity to examine long-term dynamical changes in functional architecture on an individual level. The emergence of these open-access datasets generates functional connectomic big data and simultaneously reduces the cost of data collection for researchers, which enables statistical analyses with substantially greater power and cross-center validation research with more feasibility.

**Numerous published research findings**

Over 580,000 papers related to fMRI and EEG/MEG have been published over the last eighty years (Search in PubMed using [fMRI OR “functional MRI” OR “functional magnetic resonance imaging” OR electroencephalography OR EEG OR magnetoencephalography OR MEG] as keyword), Fig. 3). The accumulating findings in this field contain valuable big data information that can be used to construct functional brain networks and explore the brain’s functional architec-

![Fig. 3. Publications related to functional brain imaging from 1941 to 2017. This figure](https://example.com/fig3.png)
activations (Dosenbach et al., 2010; Power et al., 2011), regional connectivity profiles (Cohen et al., 2008; Craddock et al., 2012; Fan et al., 2016) or clusters determined by incorporating structural and functional imaging data (Glasser et al., 2016; Shen et al., 2010b). With fine-grained functional images, a brain network size of 1 million nodes and 1 billion edges at the voxel level can be attained (Fig. 2), which provides an opportunity to ascertain the global and local topological properties of brain functional networks on a detailed level (Fig. 4A). Human brain networks derived at different spatial scales or parcellations are likely to be correlated but capture different topological mechanisms (Tohka et al., 2012; Wang et al., 2009; Zalesky et al., 2010b), which are biologically meaningful for the exploration of the functional architecture of the brain.

**Dynamic functional connectome**

Recent studies have demonstrated that the functional connectivity between regions is not constant but exhibits significant variability over a very short period (seconds or minutes) (Chang and Glover, 2010; Hutchison et al., 2013a). Dynamic functional connectivity can be estimated using sliding window correlations (Hutchison et al., 2013b; Jones et al., 2012; Kiviniemi et al., 2011; Tagliazucchi et al., 2012), Kalman filtering (Kang et al., 2011), or single volume co-activation (Liu and Duyn, 2013). Within these frameworks, researchers can generate dynamic functional connectomic big data (Allen et al., 2014; Chu et al., 2012; Zalesky et al., 2014) (Fig. 4B) and identify time-varying characteristics of brain networks, such as the core-periphery configuration of dynamically stiff regions and the network reorganization that underlies the learning progress (Bassett et al., 2011, 2013), the highly fluctuating connections among functional modules and efficient connection patterns over time (Zalesky et al., 2014), the time-varying discrete functional connectivity states across individuals (Allen et al., 2014; Calhoun et al., 2014), and the structural constraints of the dynamical functional networks (Liao et al., 2015). With an increased temporal resolution in brain activity recording derived from advanced neuroimaging and neurophysiological techniques, this big data is likely to resolve the novel time-varying organizational principles of human brain networks that cannot be captured by stationary functional connectome analyses.

**Independent component analysis**

Independent component analysis (ICA) is a data-driven blind source separation method for decomposing a multivariate signal into independent subcomponents (Hyvärinen et al., 2004). Using spatial ICA on resting-state fMRI data, researchers have successfully extracted various distinct functional components by identifying correlated spontaneous oscillations (Damoiseaux et al., 2006; Smith et al., 2009, 2015; van de Ven et al., 2004). The spatial pattern of these components corresponds to the collections of regions that are co-activated in the task paradigm (Smith et al., 2009). D) Machine learning algorithms, e.g., support vector machines, have been extensively employed in functional connectomic big data to predict the categorical or continuous variables of the testing data (Shen et al., 2010a).
et al., 2009) (Fig. 4C). In contrast to spatial ICA, temporal ICA can identify distinct ‘temporal functional modes’ based on their temporal independence with minimal spatial restriction (Biswal and Ulmer, 1999; Calhoun et al., 2001). Although temporal ICA is considerably less robust than spatial ICA in fMRI data with improved quality and increased temporal sampling rates can compensate for this shortcoming. Using fast-acquisition fMRI data (Moeller et al., 2010) and temporal ICA methods, Smith et al. (2012) discovered that the patterns of correlation and anti-correlation associated with the default mode network can be subdivided into several functionally distinct and spatially overlapping networks, which indicates that large-scale neuronal dynamics can share a substantial anatomical infrastructure. Collectively, ICA offers practical data-driven methods for mining functional connectomic big data, which enables the exploration of the patterns of spatial or temporal components in the brain’s functional architecture.

**Machine learning**

An important goal of functional connectomic big data is to reveal the neuronal pathways that underlie individual behaviors and brain diseases from a systematic perspective and identify network-based biomarkers for predicting cognitive performance, disease diagnosis, severity, and prognosis. To achieve this objective, a variety of multivariable analysis methods and machine learning algorithms have been applied to investigate functional brain networks. Traditional univariate analysis considers the features of network elements as independent variables while disregarding their potential relationship and introduces difficulties with correcting for multiple comparisons. Conversely, multivariable methods, such as multivoxel pattern analysis (MVPA), perform the analysis based on the spatial patterns of multiple signals that can overcome the limitations of low signal-to-noise ratios and strict multiple comparison corrections in univariate analysis and increase the sensitivity of feature identification from functional imaging data (Davatzikos et al., 2005; De Martino et al., 2008). Machine learning algorithms predict the categorical or continuous variables of the testing data by learning the relationship between a training set of different strategies can be employed to split the data into training and testing datasets, including k-fold cross-validation and leave-one-out cross-validation. Although these methods offer excellent prediction accuracy in classifying functional connectomic big data, biologically meaningful information, such as functional network architecture, provided by the high-dimensional models needs to be clarified.

**Statistical analysis strategies**

In the field of brain network research, the selection of statistical strategies has always been an important question, and controlling for a balance between type I error and type II error remains a challenge for researchers who test their hypotheses. Due to the strong nexus among network elements and metrics, traditional multiple comparison correction methods that are designed for independent repeated experiments, such as Bonferroni and false discovery rate corrections, seem to be too strict to reveal significant findings due to type II errors. Conversely, some novel statistical methods are in development to overcome these defects. For example, Zalesky et al. (2010a) proposed the network-based statistic (NBS) approach, which is an important technique for controlling the familywise error rate, to identify which connections that comprise a contrast or effect of interest are interconnected. Shehzad et al. (2014) proposed a promising multivariate distance matrix regression method to explore the associations between network connectivity and phenotype (e.g., age, clinical diagnosis, and behavioral performance). This method is computationally efficient, extensible and scalable for connectome-wide association studies (CWASs). A recent paper that discusses fMRI cluster correction suggests the prominent power of the nonparametric permutation test in producing nominal results for both voxelwise inference and clusterwise inference (Ekblom et al., 2016). Collectively, these newly developed or proposed statistical methods for functional connectomic big data can accelerate discovery in functional brain network architecture.

**Test-retest reliability and reproducibility of functional connectomic big data**

The test-retest reliability of the functional brain network represents whether functional architectures are stable over time. High reliability is a fundamental necessity for individual characterization and longitudinal research. Previous studies have demonstrated that the global topological properties (e.g., small-world parameters) in small functional brain networks (~50–1000 nodes) can achieve fair to good test-retest reliability, which is substantially dependent on the selection of parameters during functional connectome analysis, such as filtering frequency (Braun et al., 2012; Liang et al., 2012), removal of global signal (Braun et al., 2012; Liang et al., 2012), brain parcellation (Termenon et al., 2016; Wang et al., 2011), connectivity estimation (Liang et al., 2012; Wang et al., 2011) and network density (Braun et al., 2012; Termenon et al., 2016). Conversely, the test-retest reliabilities of high-resolution voxelwise functional networks have not been significantly investigated. One example study demonstrated that the global network metrics (e.g., small-world parameters) of high-resolution networks (~25,000 nodes) were sensitive to the scanning intervals, with fair to excellent long-term (six weeks) reliability but lower short-term reliability (20 min) (Du et al., 2013). This result may be attributed to an increase in individual tension, anxiety and fatigue during continuous short-term scanning. Other studies indicated that the reliability varied across different nodal metrics: the nodal efficiency, nodal degree and local functional connectivity density were relatively more reliable than the nodal participant coefficient metric, especially within the hub regions (Du et al., 2015; Liao et al., 2013; Tomasi et al., 2015; Zuo et al., 2012). Tomasi et al. (2016) suggested that the selection of optimal parameters, such as the network threshold R > 0.3, filtering with a bandwidth range from 0.01 to 0.08 Hz, and no global signal normalization, can produce better reliability for local functional connectivity density. Note that the scan sequence or parameters can affect the reliability of functional connectome measurements. For instance, the utilization of a spin-echo echo-planar imaging sequence can increase the sensitivity, specificity and inter-subject reproducibility in estimating functional connectivity compared with the use of gradient-echo echo-planar imaging (Khatamian et al., 2016). Longer scan duration can significantly improve the reliability of network topological properties (Birn et al., 2013; Braun et al., 2012; Cao et al., 2014a; Liao et al., 2013; Termenon et al., 2016). These studies provide experimental evidence and practical guidance for controlling intra-subject reliability in functional connectome studies.

In addition to the intra-subject variability, the test-retest reliability is also substantially reliant on the inter-subject variability. An intriguing study from Mueller et al. (2013) exhibited a heterogeneous distribution of inter-subject differences in functional connectivity across the cortex in adults: the heteromodal association cortex had significantly higher inter-subject variability, whereas the unimodal cortices had lower variability. The inter-subject variability pattern was highly correlated with the degree of evolutionary cortical expansion, which indicates a potential evolutionary root of functional variability architecture. The subsequent study from the same group set up an example for guiding individual-level brain parcellation based on inter-subject functional variability. This approach was validated by invasive cortical stimulation mapping in surgical patients, which
implies its potential usage in clinical applications (Wang et al., 2015a). This inter-subject variability pattern in functional connectomes has emerged during infancy and exhibits significant age-dependent genetic effects (Gao et al., 2014). The combination of the intra-subject reliability and inter-subject difference in the functional connectome forms a situation in which the brain network architectures are unique for each individual, which implies that the network connectivity pattern is likely to serve as a connectome fingerprint to distinguish one individual from other individuals with high accuracy either in a resting state or during a task state (Finn et al., 2015).

Few studies have focused on the reproducibility of the functional connectome findings over different data centers in past decades, which was substantially attributed to the lack of a large multisite database. Some studies that employed the multicenter datasets from the 1000 Functional Connectomes Project demonstrated a significant center effect on network connectivity patterns (Biswal et al., 2010; Tomasi and Volkow, 2011; Yan et al., 2013). Recently, Noble et al. (2016) collected resting-state fMRI data from the same subjects at eight research centers to evaluate the contributions to variability in functional connectomes due to subject, site and day-of-scan. They observed that the brain connectivity differences were predominantly obtained from subject effects rather than centers and that longer scan duration increased the reliability of connectivity estimation. The publishing of databases with multisite scans (Table 1), especially databases with multiple scan sessions for each subject, such as CORR (Zuo et al., 2013), provides a significant opportunity to assess the reliability and reproducibility of the functional connectome.

The application research of functional connectomic big data

Cognition and behavior

Functional connectomic big data has been employed to study the brain network basis that underlies a variety of cognitive functions in different settings. From a high-resolution network perspective, van den Heuvel et al. (2009) have demonstrated that the efficiency of resting-state functional brain networks was positively related to individual intelligence quotients and that the most pronounced effects were observed in several functional hubs, such as the medial prefrontal and parietal regions. Liu et al. (2016) discovered that the functional connectivity strength of voxelwise brain networks was associated with spatial working memory performance in the regions of the default-mode, visual, dorsal attention and fronto-parietal systems and that the functional hubs contributed a large proportion of the variance in the individual differences in spatial working memory (Fig. 5A). The functional architecture of brain networks can be selectively modulated across cognitive tasks. Using graph-based modularity analysis at the voxel level, a recent fMRI study demonstrated that the functional connectivity decreased within the default mode module and increased within the executive-control module as the N-back working memory task load increased, and the salience module became more strongly connected with both the default mode and the executive-control modules. These findings suggest that the integration of specialized brain networks can be dynamically reorganized in response to cognitive demands (Liang et al., 2016). Using fMRI of 64 distinct task states and resting-state fMRI data, Cole et al. (2014) demonstrated that the modulation of network connectivity patterns was generally small across tasks and was primarily shaped by the intrinsic network architecture during rest.

Mining published studies is also a promising method for revealing the relationship between the connectome and cognition and behaviors. Based on the semantic processing activation-related regions identified by a meta-analysis across 120 task-fMRI studies (Binder et al., 2009), Xu et al. (2016) established a tri-module network model that involves the default mode, the left perisylvian and the left frontoparietal networks. They suggest that the functional architecture relevant to the three networks may underlie human semantic processing for incorporating the interactions of memory, language and control. By establishing a novel, data-driven author-topic hierarchical Bayesian model in a large dataset of 10,449 published neuroimaging experiments, Yeo et al. (2015) obtained a group of cognitive components shared across 83 BrainMap-defined task categories, which was supported by a complex pattern of functionally flexible and specialized regions distributed across the association cortex. The functionally specialized regions were strongly coupled for the same components to form partially isolated brain networks, and the functionally flexible regions participated in multiple cognitive components and supported the integration of specialized brain networks (Fig. 5B). Using a similar model with 9208 imaging experiments of 77 tasks, Bertolero et al. (2015) highlighted a strong spatial correspondence between the cognitive components and the intrinsic functional modules and demonstrated that the network connectors that link different modules were primarily distributed in the frontal and parietal association cortex and were involved in multiple cognitive functions. Recently, Smith et al. (2015) demonstrated a prominent example in the statistical exploration of functional connectomic big data. To analyze the relationship between the brain’s functional architecture and individual phenotypic measures, they employed resting-state fMRI data and ICA to identify 200 brain nodes and perform canonical correlation analysis (CCA) to correlate these nodes with 158 behavioral and demographic measures. They highlighted a strong mode of subject covariation that was spread along a single ‘positive-negative’ axis associated with demographic, psychometric and lifestyle measures, as well as specific brain connectivity patterns in the default mode and attention networks (Fig. 5C). These studies provide excellent examples regarding how functional connectomic big data contributes to the exploration of the relationship between the brain’s functional architecture and cognition and behavior.

Development and aging

Functional connectomic big data has the natural advantage of including large samples to delineate the trajectory of the changes in functional architecture from the neonatal stage to aged stages. However, studies are separately focused on specific age stages of a person’s entire life, including the fetus, infancy, childhood, adolescence and adulthood. At the neonatal or infancy stages, the brain’s functional networks have been demonstrated to exhibit small-world configurations, modular structure, and rich-club architectures (Cao et al., 2016; Fransson et al., 2011; Scheinost et al., 2016; van den Heuvel et al., 2015). Using resting-state fMRI data and voxel-based graph-theory analysis in infants aged approximately 31–42 postmenstrual weeks, Cao et al. (2016) revealed that the functional connectivity strength and heterogeneity significantly increased in the primary motor, somatosensory, visual, and auditory regions but exhibited a lesser increase in high-order default-mode and executive-control systems. The core structures of hubs and rich clubs in primary regions had emerged at approximately 31 postmenstrual weeks and expanded with age, accompanied by increased local clustering and the shortest path length, which indicates a transition from a relatively random configuration to a more organized configuration. Based on a large sample of more than 400 babies, Gao et al. (2014) revealed that during the first two years after birth, long-range connections of hub regions were increased to form an efficient organization of the functional brain networks. Inter-subject variability in functional connectivity profiles revealed an adult-like distribution over the brain and underwent a U-shaped growth trajectory, which may suggest “skill learning” development during the infancy stage. Large sample data collection from childhood to adulthood is relatively easier than data collection for babies and infants, and the majority of recent studies that address this stage included hundreds to a thousand scans. These studies demonstrated that the hub distributions of the functional networks stabilized in this stage and
were predominantly located in the default-mode regions, the visual cortex, the insula and the striatum (Cao et al., 2014b; Zuo et al., 2012). The regional centralities of the association cortex increased with age, whereas the regional centralities of the primary cortex decreased with age (Sato et al., 2014, 2015; Zuo et al., 2012). Different functional systems of the whole-brain network exhibited diverse trajectories with development (Gu et al., 2015). Based on graph theory and multivariate machine learning approaches, the connectivity pattern within the whole-brain connectomic big data can be used to successfully predict maturation (Dosenbach et al., 2010) or brain age (Cao et al., 2016; Qin et al., 2012). Throughout a broader lifespan from 7 to 90 years, studies have shown that increasing age is accompanied by linear decreasing segregation of brain systems, as indicated by the modular structure of the functional brain networks (Cao et al., 2014b; Chan et al., 2014). Inverted-U shaped trajectories were observed for the local efficiency and rich club architecture, and the trajectories of long- and short-distance functional connections were inverse, which may indicate the reorganization of connectivity concentration and distribution in the brain’s functional architecture during a lifespan (Cao et al., 2014b). Strong divergence in network changes across brain regions were observed. Using the large sample datasets from 1000 Functional Connectomes Project, Biswal et al. (2010) demonstrated that the age-dependent changes in functional connectivity were heterogeneous in regions of the default mode network. Tomasi and Volkow (2012) revealed that aging was associated with long-range functional connectivity density decreases in the default mode and dorsal attention networks and increases in the somatosensory and subcortical networks. These connectomic big data studies significantly improve our understanding of the reorganization in human functional architecture during a lifespan and provide valuable information for guiding developmentally informed policies and treatment for populations from the fetus to the aged.

### Brain disorders

The application of functional connectomic big data becomes increasingly important when either studying pathophysiology or exploring the network-based biomarkers of brain disorders, and different data features characterize different aspects of neuropsychiatric disease. High-resolution functional networks with graph theoretical frameworks have been extensively employed to identify detailed network abnormalities associated with various brain disorders, including Alzheimer’s disease (AD) (Dai et al., 2015), attention-deficit/hyperactivity disorder (Di Martino et al., 2013), Parkinson’s disease (Gotlib et al., 2013), schizophrenia (Skudlarski et al., 2010; Wang et al., 2014b), major depressive disorder (Guo et al., 2015; Wang et al., 2014a), social anxiety disorder (Liu et al., 2015), obsessive-compulsive disorder (Hou et al., 2014) and amyotrophic lateral sclerosis (Zhou et al., 2016). Dai et al. (2015) demonstrated that the densely connected brain hubs with long-range connections were highly targeted and significantly disrupted in AD, which provided connectomic insights into the pathophysiological mechanisms of network dysfunction in this...
High-resolution networks can provide potential connectome biomarkers for early diagnosis (Dai et al., 2012; Li et al., 2016) and possess potential for evaluating drug and stimulus treatments of psychiatric disorders, such as depression (Perrin et al., 2012; Wang et al., 2015d) (Fig. 6A). Combining multimodal imaging and biological information is also important for revealing the pathology of disease from a multidimensional perspective. A recent study collected over 7700 multimodal imaging datasets, including fMRI, sMRI, ASL and PET, from the ADNI database and performed multifactorial data-driven analysis. The results revealed that intra-brain vascular dysregulation is an early pathological event during disease development and suggested a pathological model that characterizes the development of different biomarkers during disease development (Iturria-Medina et al., 2016) (Fig. 6B). Novel meta-analytic approaches, in which network analysis methods are applied to previously published studies, demonstrated that the structural abnormalities of 26 brain disorders were primarily concentrated in the highly connected brain hubs of either the structural or functional co-activation networks (Crossley et al., 2014). D) Using canonical correlation analysis and clustering algorithms in large sample data, depressive patients could be clustered into four neurophysiological subtypes according to distinct patterns of dysfunctional connectivity in the limbic and frontostriatal networks (Drysdale et al., 2017).

**Fig. 6.** Exploring functional connectomic big data for brain disorders. A) A high-resolution network analysis revealed that the functional connectivity strength (FCS) in the bilateral dorsal medial prefrontal cortex (dmPFC) and bilateral hippocampus can be modulated by eight weeks of treatment with a selective serotonin reuptake inhibitor—escitalopram—in patients with major depressive disorder (MDD) (upper left), and the changes of FCS in dmPFC are significantly associated with the changes in clinical symptoms (upper right). A post hoc analysis revealed that the FCS in the bilateral dmPFC in the MDD patients is significantly higher than the FCS in the bilateral dmPFC in the HC group at the baseline and is significantly reduced in patients after treatment, whereas the FCS in the bilateral hippocampus shows the opposite pattern (bottom) (Wang et al., 2015b). B) A study using collection of multimodal imaging data and biological information with a large sample size in Alzheimer’s disease revealed that intracerebral vascular dysregulation is an early pathological event and suggested a pathological model that characterizes the development of different biomarkers during disease development (Iturria-Medina et al., 2016). C) Novel meta-analytic approaches, in which network analysis methods are applied to previously published studies, demonstrated that the structural abnormalities of 26 brain disorders were primarily concentrated in the highly connected brain hubs of either the structural or functional co-activation networks (Crossley et al., 2014). D) Using canonical correlation analysis and clustering algorithms in large sample data, depressive patients could be clustered into four neurophysiological subtypes according to distinct patterns of dysfunctional connectivity in the limbic and frontostriatal networks (Drysdale et al., 2017).
9874 patients and 11,502 healthy controls, Crossley et al. (2014) demonstrated that the structural abnormalities of 26 brain disorders were primarily concentrated in the highly connected brain hubs of either the structural network or functional co-activation networks (Fig. 6C). In a subsequent study, Crossley et al. (2015) conducted a meta-analysis of 168 voxel-based morphometry studies of 14 neurological and 10 psychiatric disorders. They exhibited distinct affected patterns for the two classes of disorders: the neurological disorder appeared to significantly affect the sensorimotor and frontoparietal networks, whereas psychiatric disorders significantly affected the visual and default mode networks. The recent study of these authors, based on 314 task-related functional neuroimaging studies and including more than 5000 patients with schizophrenia and over 5000 controls, highlighted that the regions with abnormal functional activations in schizophrenia were primarily located in the hub nodes of the normative connectome (Crossley et al., 2016). Kaiser et al. (2015) investigated large-scale network dysfunction in major depressive disorder by performing a meta-analysis of 25 functional connectivity studies, including over 1000 participants. The results indicated that depression was characterized by hypoconnectivity within the frontoparietal network and between the frontoparietal and dorsal attention networks and hyperconnectivity within the default mode network and between the default mode and frontoparietal networks, which provides an empirical foundation for a network dysfunction model that explains the core cognitive and affective abnormalities in depression. Collectively, taking advantage of different approaches to functional connectomic big data can either provide general concepts for disease pathology or identify potential biomarkers for the early prevention, diagnosis, prognosis and treatment of diseases.

Future directions

As previously described, the emergence of functional connectomic big data and the advancement of data mining techniques provide unprecedented opportunities to understand the functional architectures of the human brain. However, many challenging issues in the functional connectomic big data era need to be urgently addressed in the near future.

Big data versus small sample

In the framework of functional connectomic big data, one of the typical features is a large quantity of samples, which is crucial for reliably identifying the brain’s functional architecture due to improved statistical power and to enable the generalization of findings across populations. The data-driven analysis approaches in a large sample can obtain novel findings regarding the functional architecture of the brain, which cannot be obtained by small hypothesis-driven studies. However, the case that “more is better” is not always valid. In contrast with small datasets, functional connectomic big data are easily polluted by noise during research design and data collection. For example, a large sample size often increases the heterogeneity of the data due to the diversity of individual demographics and behavioral performance. The collection of big data usually involves multiple different imaging scanners and protocols, which are likely to have a significant influence on brain data analyses. Thus, the emergence of big data cannot eliminate the necessity of the existence of small sample studies since the latter can collect a homogeneous population and is more flexible in experimental design and control. Small sample studies may be better suited to answering specific, hypothesis-driven questions than larger heterogeneous studies, in which the effect is averaged. Note that the relationship between big data and small sample studies is not contradictory; rather, they are complementary. The data-driven findings from functional connectomic big data can be considered to be prior knowledge to guide future small sample studies.

The structural substrates of the functional connectome

Elucidating the structural substrates of the brain's functional architecture is an important and challenging issue. Several studies have demonstrated a significant association between functional and structural connectivity across cortical regions (Hagmann et al., 2008; Honey et al., 2009). However, the network topological properties, such as the modularity and nodal centralities, between the two modalities in one-to-one comparisons differed (Alexander-Bloch et al., 2013; Baria et al., 2013; Park et al., 2008; Wang et al., 2015e). Despite this finding, researchers have tended to link different properties in the two types of networks to illustrate the possible organizational principle of the brain. For example, van den Heuvel and Sporns (2013) demonstrated that the cortical rich club of the structural networks serves a central role in the cross-linking of distinct modules of the human functional networks. Several models have been constructed to predict both the strengths and the spatial patterns of the functional networks based on anatomical constraints (Goni et al., 2014; Shen et al., 2012; Sporns and Kotter, 2004; Vertes et al., 2012). Note that several commonly employed white-matter properties estimated from diffusion MRI (e.g., fractional anisotropy) can only provide approximate reflections of specific microstructural attributes, such as fiber organization, axon density and myelination. Given that the accurate biological interpretation for these indices remains ambiguous and controversial (Jones et al., 2013), the relevant results should be carefully explained to avoid overestimation. Recently developed diffusion MR models, such as CHARMED (Assaf et al., 2004), and other MR contrast mechanisms (Sled and Pike, 2001), are expected to provide more accurate microstructural properties. Cytoarchitecture describes the spatial distribution pattern of the neuronal cell bodies of brain regions at the micro-level (Amunts and Zilles, 2015); however, its relationship to the functional brain network remains substantially unknown. Using promising microtome and digitalization techniques, Amunts et al. (2013) created an ultra-high-resolution three-dimensional model of a human brain at the nearly cellular resolution of 20 μm. The future development of techniques to link the fine-grained functional network big data to these macro- and microscopic structural properties and to establish the proper model for illustrating their interactions is important.

The electrophysiological and metabolic basis of functional connectome

In BOLD fMRI-based connectome studies, functional connectivity is typically estimated from the correlations between activity time courses of anatomically isolated regions. However, the exact neurophysiological and biochemical substrates that underlie functional connectivity remain substantially unclear (Florin et al., 2015; Logothetis, 2008). Using simultaneous EEG/fMRI recording, local field potential (LFP)-fMRI, or LFP-optical recording techniques in animals, recent studies have demonstrated that the oscillation of BOLD functional connectivity is related to synchronized neuronal electrophysiological activity (Niessing et al., 2005; Shmuel and Leopold, 2008), specifically the gamma frequency band (Logothetis et al., 2001; Niessing et al., 2005; Shmuel and Leopold, 2008), which is suggested to be strictly coupled with the periodic inhibition of pyramidal cells (Jonas et al., 2004). Based on sequential recording of ECoG and fMRI with equivalent task paradigms in humans, the local BOLD fluctuations are highly associated with the properties of ECoG gamma frequency bands in different task situations; this finding is consistent with the finding in animals (Conner et al., 2011; Kunii et al., 2013). A few studies conducted simultaneous intracranial EEG and fMRI recording to study BOLD changes related to interictal epileptiform discharges (Murta et al., 2016; Vulliemoz et al., 2011). One recent study demonstrated that a sharp wave width rather than other features of interictal epileptiform discharges can explain a significant amount of variance in BOLD changes, which suggests that the amplitude of the BOLD signal is substantially dependent on the
duration of the underlying field potential (Murta et al., 2016). Although the sample size in the majority of the studies in this field are small due to the experimental difficulty in combining the collection of electrophysiological and neuroimaging data, especially in simultaneous recording, these findings established a possible link between BOLD fluctuations and neural activity. However, the electrophysiological and biochemical substrates that underlie large-scale functional architecture remain largely unknown. Future studies that combine electrophysiology and neuroimaging with relevant data sharing projects in this field may facilitate the accumulation of experimental data and provide important opportunities to explore reliable associations between BOLD signal and neural activity and expand the understanding of the electrophysiological substrates of functional architecture.

The highly connected hub regions of large-scale functional brain networks have particularly high metabolic demands. Using multimodal BOLD fMRI and arterial spin labeling perfusion MRI data, Liang et al. (2013) demonstrated significantly positive correlations between the nodal centralities of the voxelwise whole-brain functional networks and regional cerebral blood flow, especially in the long-range functional hubs. This relationship appeared to be significantly strengthened with increasing working memory task loads. This study also demonstrated a strict coupling between BOLD functional connectivity strength and PET metabolic measurements (e.g., CMRGlu and CMRO2). Another study based on human fMRI and PET data demonstrated that dense functional connectivity of network nodes was associated with nonlinear increases in the cerebral metabolic rate of glucose (Tomasi et al., 2013). The functional connectivity strength of network nodes was closely associated with the cortical excitatory-inhibitory ratio in both human and macaque brains, which indicates the relationship between the microscale chemoarchitecture and the topology of functional brain networks (van den Heuvel et al., 2016). These findings imply that the large-scale functional networks constructed from BOLD fMRI data may partially capture the electrophysiological and metabolic signatures. Future studies that combine multimodal electrophysiological recording (e.g., EEG/ECOG) and imaging data (e.g., PET/ASL) need to be conducted to characterize the electrophysiological and metabolic substrates that underlie the functional architectures derived from functional connectomic big data.

The association between macroscale and microscale functional connectomes

Functional brain networks can be described at different spatial scales. At the microscale, the neurons communicate with each other by releasing and receiving neurotransmitters or by passing electric currents via synaptic connections to form a vast neuronal network. The structures of these networks can be detected using high-resolution anatomical methods (Amunts et al., 2013; Amunts and Zilles, 2015), and the neural activity can be recorded with optical or electrical techniques (Engelbrecht et al., 2014; Palmini, 2006). When a cohort of neurons is active, they cause changes in local field potentials, hemodynamics, biochemistry and metabolism, which can be observed with macroscale imaging methods. Although slightly inconsistent, the functional brain networks at the macroscale partly reflect the distant communication of clustered neurons at the microscale (Leichtman et al., 2014; Scholtens et al., 2014). However, functional connectomes at different scales capture different biological significance. For example, functional networks at the macroscale can describe the network architecture of the whole brain in vivo in terms of the communication among functionally independent brain regions or different functional systems. Within this framework, various psychological and disease models have been proposed to delineate biological mechanisms that underlie human behavior (e.g., top-down control) (Dosenbach et al., 2008; Petersen and Sporns, 2013) and brain diseases (Iturria-Medina et al., 2016; Zhou et al., 2012). In contrast, current microscale brain networks can only be constructed within a local area due to technical limitations; however, they are able to describe the organizational and dynamical principles of the neurons (Schroeter et al., 2015; Shimono and Beggs, 2015). Studies at both the micro-level and the macro-level are important to heighten our understanding of functional architectures of the human brain at different scales and how neural spikes at the micro-level facilitate communication among functional systems at the macro-level. These studies provide a wealth of information to bridge the mechanistic gap between different scales and to establish and validate biophysical models that shape macroscale functional brain networks.

The development of novel analysis strategies for functional connectomic big data

Developing novel analytical strategies for functional connectomic big data is a fundamental requirement toward pure research on the brain’s functional architecture and various applied studies in this field.

First, improvements in the temporal resolution of functional data can measure abundant information regarding time-varying functional architecture over shorter time periods (Allen et al., 2014; Liao et al., 2013). The field of dynamic brain networks is rapidly growing; however, many challenging issues exist, including: i) establishing computational models that capture spatial and temporal coupling across both macro- and microscales, ii) validating the reproducibility of developed and developing dynamic analysis methods and metrics using large datasets, and iii) specifying the features of dynamic networks for predicting different cognitive behaviors and different disease states. Several recent studies suggested that the time-varying functional connectivity during rest may be associated with sampling variability and head motion (Laumann et al., 2016), as well as physiological signals, including arterial CO2 and heart rate/heart rate variability (Nikolaou et al., 2016). Future studies are required to clarify the biological mechanisms that underlie the dynamic functional connectome and develop methods to extract neural dynamic information from the confounding physiological noise.

Second, functional connectomic big data are characterized by high-dimension information that captures the organizational principles from local nodes and connections to systems and global architectures. Combined with other types of information, including anatomic, genetic, demographic, cognitive, and biophysical records, the amount of data is unprecedentedly vast. The field requires additional effort toward developing and introducing multivariate analyses, such as machine learning approaches, for data mining. Various deep learning architectures, such as deep neural networks, convolutional deep neural networks and recurrent neural networks, have demonstrated their powerful capabilities in various tasks in computer vision, audio recognition, automatic speech recognition and natural language processing (Ciregan et al., 2012; Dahl et al., 2012; Hinton et al., 2012). Note that a limited training sample can easily cause overfitting in deep learning models; thus, this method has been narrowly employed in functional brain network analysis. Developing optimized models and other corresponding algorithms that are suitable for connectomic big data may help to identify brain network architectures related to genes, cognition and diseases.

Finally, several intriguing studies have demonstrated the significant power of network-based statistical analysis strategies, such as NBS and multivariate matrix distance regression methods (Kim et al., 2014; Zalesky et al., 2010a); however, statistical approaches that are designed for functional connectomic big data analysis remain lacking. In contrast to traditional statistics, network-based statistics address the problem that the network elements are highly interactive across regions and connections and across network features and densities. With an exponential growth in network data, the development of novel network-based statistical approaches, especially correction methods for multiple comparisons, is urgently required to control for both type I and type II errors.
Reconsidering brain diseases with functional connectomic big data

As an important output of functional brain network studies, the significance of mining in big data is either theoretically or methodologically important for brain diseases. A variety of studies have demonstrated alterations of functional architecture in brain networks in various neuropsychiatric disorders (Bullmore and Sporns, 2009; Fornito et al., 2015; Stam, 2014; Xia and He, 2011). These findings have provided a critical conceptual understanding of disease pathophysiology; however, many inconsistent results are obtained across studies due to the lack of statistical power in small samples or the selection of different analysis strategies (e.g., Gong and He, 2015). Functional connectomic big data provides significant opportunities to examine the reproducibility of network dysfunctions in brain disorders, especially to reliably specify disease-specific alterations in functional architectures and facilitate the accumulation of knowledge about disease pathologies (Castellanos et al., 2013).

Functional connectomic big data serves a practical role in clinical diagnosis and treatment evaluation. For many brain disorders, the ground truth for distinguishing patients from healthy individuals is established based on clinical diagnoses that involve symptom estimation and neuropsychological tests. Although the criteria for the assessment of mental disorders are identical according to international diagnosis standards, the diagnostic results regarding the type of illness may substantially depend on the experiences of clinicians. The job is often difficult, especially given the notion that some mental illnesses form continuous spectra. Therefore, the potential contribution from functional connectomic big data is to reconsider the diagnostic delineations between different brain disorders and the subtype classifications within each type of disorder (Insel and Cuthbert, 2015). Recently, an intriguing study from Drysdale et al. (2017) offered a promising example for defining subtypes of depression from big imaging data. Using R-fMRI data from 1200 subjects, they discovered that the depressive patients were clustered into four neurophysiological subtypes according to distinct patterns of dysfunctional connectivity in the limbic and frontostriatal networks. This division can accurately predict responsiveness to transcranial magnetic stimulation therapy (Fig. 6D). Many imaging studies have revealed significant associations between features of functional networks and patients’ symptoms, which provide valuable information regarding biomarker research for the diagnosis or treatment of diseases (Fornito et al., 2015; Gong and He, 2015; Stam, 2014; Xia and He, 2011). However, note that the majority of these studies were based on relatively small sample sizes; future studies that employ functional connectomic big data, especially from multiple centers, can provide solid validation for this biomarker research (Arbabshirani et al., 2017). Modern medicine tends to collect comprehensive information from patients, including genetic, behavioralfamilial, biomechanical and biophysical data, as well as neuroimaging and neurophysiological data. A promising study by Finn et al. (2015) indicated that functional network architecture derived from fMRI data possibly holds an individual fingerprint signature. By combining genomic, phenotypic and functional connectome measures, establishing an individual fingerprint that can reliably identify patients with brain disorders from healthy people and optimizing individual treatment plans may be possible in the future. These crucial concepts and research are expected to facilitate the progress of personalized medicine.

Establishing computational platforms for functional connectomic big data

Currently, the design of new frameworks to handle the storage, sharing, computation and visualization of functional connectomic big data is urgently required. Several powerful data management systems have been developed, including the Extensible Neuroimaging Archive Toolkit (XNAT) (Marcus et al., 2007), the Laboratory of Neuro Imaging Image Data Archive (LONI IDA) (Van Horn and Toga, 2009), the Longitudinal Online Research and Imaging System (LORIS) (Das et al., 2011) and the Combined Online Information System (COINS) (Scott et al., 2011). Several available toolboxes have been employed for the analysis and visualization of functional connectomes, including the GReph thEoreTical Network Analysis (GRETNA) toolbox (Wang et al., 2015c), the Brain Connectivity Toolbox (BCT) (Rubinov and Sporns, 2010), the CONN (Whitfield-Gabrieli and Nieto Castanon, 2012) and the BrainNet Viewer (Xia et al., 2013). The GRETNA toolbox enables the construction and topological analysis of networks based on MATLAB codes with parallel acceleration from PSOM. The BrainNet Viewer provides flexible functions for the visualization of brain networks. However, these toolboxes were designed for networks of relatively small size (with 1000 nodes) without optimization for the big data in functional brain networks (more than 100,000 nodes). One recent study proposed a hybrid GPU-CPU framework to accelerate the computation of high-resolution voxel-based brain network analysis (Wang et al., 2013). The study set an inspiring example that indicates that the introduction of leading-edge computer science techniques to the field would provide operative solutions for computation problems associated with connectomic big data. In the connectomic big data era, the development of high performance computational platforms with powerful computational ability and visualization functions are important to accelerate discovery of the brain’s connectivity architecture.

Acknowledgement

We would like to thank the three reviewers for their constructive comments and suggestions. This work was supported by the Natural Science Foundation of China (Grant nos. 81401479, 81627167, 8162018016, and 31521063), the Beijing Natural Science Foundation (Grant nos. Z15110003915082, Z16110000216125, and Z16110000216152), and Changjiang Scholar Professorship Award (Award no. T2015027). The authors declare no competing interests.

References


