MACHINE LEARNING APPROACHES TOWARDS HOLISTIC BRAIN FUNCTIONAL SPACE DISCOVERY FROM FMRI BIG DATA

by

XIANG LI

(Under the Direction of Tianming Liu)

ABSTRACT

While functional neuroimaging has been going through significant advancement in the past decade, there remains a fundamental question of how we can utilize the imaging data to describe brain functional behavior in a reproducible and faithful manner. In this dissertation I have elucidated a series of my works all aims at answering the above question, yet from three different perspectives both regarding to the neuroscience implication and to the scale of the data. Firstly, statistical models are built to characterize the changes of functional organization pattern in individual brains (small size), in order to detect the quasi-stable brain states. Secondly, the concept and corresponding framework of functional connectomics summarize the common connectivity patterns within a group of individuals (medium size), and use them for dynamic transition modeling. Thirdly, dictionary learning method and its distributed implementation enable us for the efficient functional network discovery from population-wise data (large size). Based on these works, we could then eventually learn the set of holistic brain functional space from fMRI big data, through which individual signals can be effectively encoded and analyzed.

INDEX WORDS: big data, distributed computation, machine learning, fMRI

MACHINE LEARNING APPROACHES TOWARDS HOLISTIC BRAIN FUNCTIONAL SPACE DISCOVERY FROM FMRI BIG DATA

by

XIANG LI

B.A., Shanghai Jiaotong University, China, 2001

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2016

MACHINE LEARNING APPROACHES TOWARDS HOLISTIC BRAIN FUNCTIONAL SPACE DISCOVERY FROM FMRI BIG DATA

by

XIANG LI

Major Professor: Tianming Liu

Committee: L. Stephen Miller
Kang Li
Xiaoping Hu

Electronic Version Approved:
Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2016

DEDICATION

This work is dedicated to the endeavors in discovering knowledges and revealing hidden patterns from the vast treasure of the functional neuroimaging. It is the vast uncertainties and grand challenges in this field that makes all the studies towards it meaningful and interesting.

ACKNOWLEDGEMENTS

First of all, I would like to address that this work, along with all the related projects and publications, cannot be achieved without the help from my advisor, Dr. Tianming Liu. His vision in the advancement of functional neuroimaging analysis, plus his enthusiasm in performing real important (albeit difficult) researches, have been guiding me from the very first moment in my research career. It has always been a pleasure in being part of the young yet highly proficient (and passionate) team of researchers Dr. Liu has established during these few years, including Dr. Kaiming Li, Dr. Dajiang Zhu, Dr. Xi Jiang, Dr. Hanbo Chen, Dr. Jinglei Lv and many others. The team working spirit across our groups is the key in all the works we have done. On a broader scale, I would like to appreciate the mentorship from all of the collaborators during my PhD study. The collaboration projects with Dr. Jieping Ye, Dr. Hanchuang Peng and Dr. Jing Zhang have constituted several major parts of this work.

Moreover, I would like to thank Dr. L. Stephen Miller, Dr. Kang Li, and Dr. Xiaoping Hu for joining my dissertation committee. Their insightful suggestions and encouragements are most helpful for my dissertation work.

Finally, I want to thank my family in supporting me to pursuit my PhD studies. Their trust and help is invaluable through these 6 years of my life, in all the aspects.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

1.1 Thesis Statements

The work in this thesis is focused on one single objective: to better characterize the functional organization pattern and the cognitive process of human brain, i.e. "how brain works". A series of studies from me and our group have been performed towards this goal, by developing models on the functional Magnetic Resonance Imaging (fMRI) data on various scales and conditions. Specifically, I have developed mathematical models for characterizing the functional network dynamics and the functional-structural relationships of the brain, including fiber-guided functional connectivity modeling [1], sliding time window-based functional connectomics analysis framework [2-7], the two-level MCMC model for simultaneous temporal and spatial pattern inference [8], and sparsity-regularized dictionary learning method for functional network decomposition [9-11].

Further, as the availability and heterogeneity of the neuroimaging data keeps growing, it is becoming much more important to utilize population-level data for learning the holistic brain functional networks space, rather than relying on the dominant features from limited number of subjects. As highlighted in [12], the need for large-scale fMRI analysis is especially imminent to overcome the bias and false-positives in traditional hypothesis-based studies. In addition, fMRI big data posed grand challenges on the analysis methods: data size quickly out-grows the memory capacity and computational power. To address such need for large-scale and fast fMRI analytic methodologies, my analysis of fMRI data

has been expanded to provide solutions for the big data analytics problem in neuroimaging researches by developing fast and parallel machine learning framework. Such solutions have become more challenging and in high-demand. The framework incorporates integrated informatics system and the fast and scalable algorithms for high-throughput neuroimaging researches.

1.2 Contributions

Concepts and Methods for Dynamic Functional Connectomics: Motivated by the discovery both from my own work [1] and other literature reports [13], it has been recognized in my research that the functional brain activation is dynamic and modulated by state-by-state pattern. Based on such observation I developed the sliding time-window based functional connectivity modeling of fMRI data, then extent it into group-wise dynamic functional connectomics study based on the DICCCOL system [14]. The connectomics is defined by the most representative and/or discriminative connectivity features learned from aggregated individual functional dynamics using sparse representation. Results from various applications of the connectomics modeling show that it can effectively characterize the functional brain behavior using a much reduced set of brain states (i.e. encoding) [3]. Further, the state transitions modeling based on the connectomics characterization has bene shown to be a powerful approach to model the cognitive process as Markov process [6]. The dynamic functional network analysis framework and the resulting characterization of common/signature functional connectomics from healthy subjects and patients with mental disorders has been applied

by various of our collaborators including labs in Yale University, Zhejiang University as well as two top hospitals in China.

Dictionary Learning Methods for Functional Network Discovery: Based on the success in applying sparse representation for learning from functional connectivity, we further applied the similar sparsity-regularized matrix decomposition method on the raw fMRI data in order to discovery meaningful components (i.e. functional networks). The dictionary learning method is similar to the Independent Component Analysis (ICA) which has been widely applied on fMRI analysis [15, 16], yet is shown to be more advantageous over ICA especially for its capability in identifying overlapping functional networks and their interactions. Such discovery leads us to the development for the Holistic Atlases of Functional Networks and Interactions (HAFNI) [10] system which pooled the functional networks from the whole Human Connectome Project (HCP) Q1 database [17] and manually aligned them to form an atlas. In addition, the dictionary learning method was the core component for my advisor Tianming Liu in one of his NSF grant 1439051 for the study of functional architecture of brain.

Distributed Rank-1 Dictionary Learning (D-r1DL) framework: The D-r1DL framework is developed based on the r1DL model for functional network discovery as introduced above, with the feature of enabling distributed and high performance computation. The framework has been deployed on the high performance computation platforms provided by the Georgia Advanced Computing Resource Center, as well as the Amazon Elastic Compute Cloud. The current software packages implemented in PySpark for functional network decomposition through dictionary learning has the capability of handling Terabyte-level fMRI data by parallelization and in-memory data abstraction. In

addition, the algorithm has been deployed as a web service on an established neuroinformatics system (bd.hafni.cs.uga.edu) for fast individual-level (100~500 Mb) data analysis and real-time visualization.

1.3 Thesis Outline

As illustrated in Fig.1, the work of this thesis contains three highly inter-related chapters (chapter 2 to chapter 4) which focuses on functional neuroimaging data analysis from individual level (small data, more detailed) to population level (big data, more holistic). In chapter 2, the framework developed for characterizing and modeling individual functional network dynamics will be introduced, followed by the definition of "functional brain states" which is the key concept of this thesis. In chapter3 I will describe the modeling of group-wise functional connectomics and its applications. By encoding the functional brain states by the discrete connectomics, we can then efficiently characterize the process of functional brain dynamics under various conditions. In chapter 4 we will face the challenge (and the opportunity) coming from the population-wise big functional imaging data, where novel distributed, cloud-based computation solutions are provided to support the connectomics analysis. It is envisioned, and partially achieved, in this thesis that the functional space discovered from the population-wise data could be serving as "the Standard" for all the future functional imaging analysis.

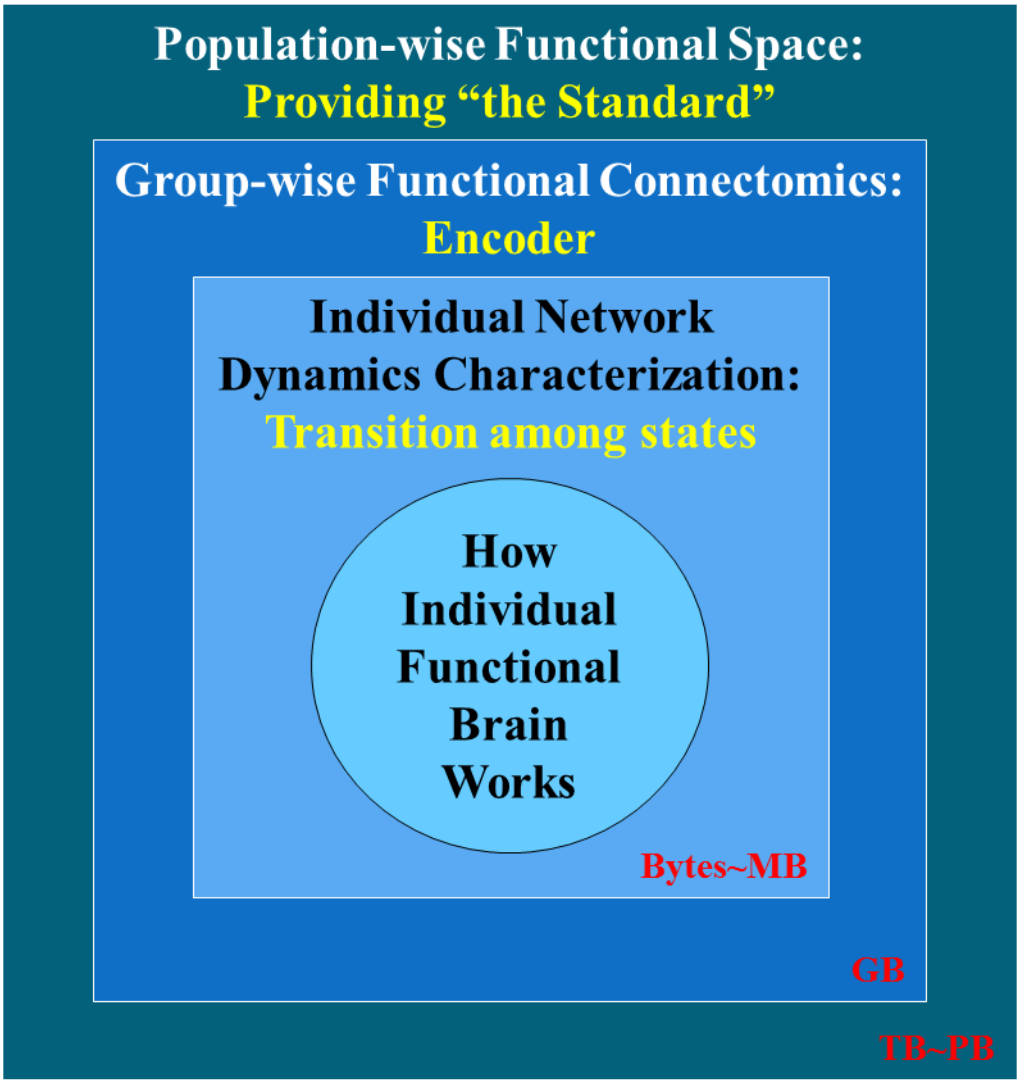

Fig.1. An illustrative overview of the three main components in this thesis from the data size perspective, which will be detailed in Chapter 2-4.

CHAPTER 2

CHARACTERIZATION OF FUNCTIONAL BRAIN DYNAMICS

2.1 Pilot Investigations and Observations on Functional Brain Dynamics

In previous researches on functional brain imaging study, especially for the connectivity analysis, a common assumption used is the temporal stationarity. Functional connectivity is computed over the entire fMRI scan and used to characterize the strengths of connections across regions [18-21]. However, accumulating literature evidence [13, 22, 23], including our own studies [24] have shown that functional connectivity is under dynamic changes at different time scales. In particular, extensive neuroscience research suggests that the function of any area of the cortex is subject to top-down influences of attention, expectation, and perceptual task [25, 26]. For instance, each cortical area runs different "programs" according to the context and to the current perceptual requirements, and dynamic functional interactions between structural connections mediate the moment-by-moment functional switching in the brain [25]. Even in the resting state, functional brain connectivity is still under dynamic changes within time scales of seconds to minutes [13]. In literature, there have been a variety of studies that investigate the problem of temporal brain state changes from different perspectives. For instance, from the fMRI blood-oxygen-level dependence (BOLD) signal processing perspective, statistical signal processing methods have been applied on fMRI signals to detect BOLD signal state change in response to stimulus (e.g., [22, 23]), and these results have been correlated to brain state change. From the brain network perspective, functional networks have been reported to form and

disappear during certain tasks, and the temporal clustering analysis (TCA) approach has been developed to detect the dynamic behavior of brain states (e.g., [27, 28]). Signal propagation from changing networks within rats and human brains was discussed in [29].

In response to the need of a more direct approach for characterizing the functional brain dynamics, my pilot study developed the model for determining functional brain state change points, by identifying abrupt alterations of functional connectivity in large-scale brain networks. Our rationale is that the brain function is integrated via large-scale structural and functional connectivity (e.g., [30-33]), and that sudden change of global functional brain connectivity is a meaningful and effective indicator of functional brain state switch. Therefore, in this model the functional brain state is defined as the specific organizational pattern of the brain's global functional connectivity [34], and brain state changes are supposed to reflect the brain's functional interaction dynamics in response to external/internal stimulus and/or previous brain states. The model is based on the fiber-centered approach to define functional connectivity on DTI-derived white matter fibers, with the basic premise that axonal fibers obtained from DTI tractography are the structural substrates of functional connectivity between brain regions [31, 35, 36], thus provide a natural anatomical localization for inference of functional connectivity. In our approach, the functional connectivity is defined as the temporal correlation between spatially remote fMRI signals extracted from gray matter voxels on the two terminals of a DTI-derived axonal fiber. That is, we measure the temporal correlation of fMRI time series of two ends of a fiber to define the functional connectivity between the gray matter voxels that it connects. The functional connectivity patterns of all of the DTI-derived white matter fibers within the whole brain are then concatenated into a descriptive functional feature vector to

represent the brain's state, called functional connectivity vector (*FCV*), as illustrated below in Fig. 2.

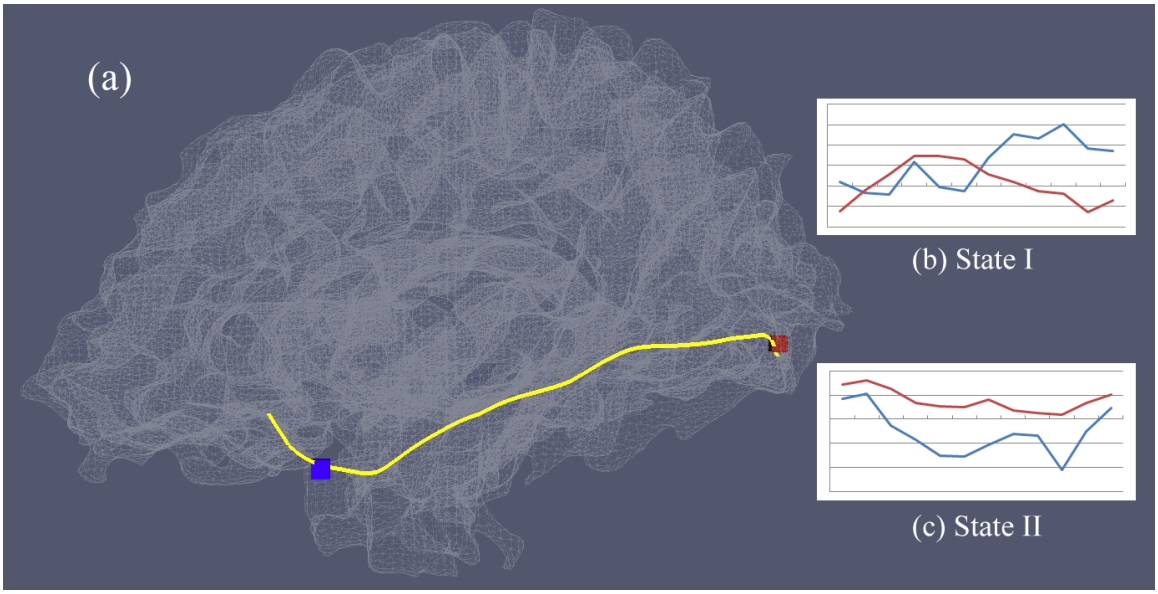

Fig.2. (a) Example of a SCGM voxel pair shown in red and blue boxes that are connected by DTI-derived fibers (in yellow). (b) The fMRI time series from the two voxels have low correlation within a specific time window (State I). (c) The fMRI time series from these two voxels are relatively higher correlated within another time window.

The functional brain state change points are then determined by the abrupt changes of the *FCV* patterns calculated by the sliding window approach along the time series. Specifically, to quantitatively characterize the functional connectivity dynamics, we define the *FC* value on axonal fibers between voxel pair [v_g, v_h] in time window [t_i, t_j]:

$$FC(v_g, v_h, t_i, t_j) = Pearson\ correlation\ between\ fMRI\ signals\ of\ v_g, v_h\ from\ t_i\ \text{to}\ t_j \quad (1)$$

Assume that the totally scan length is *l* time points, and the time window size is *s*, we could apply a sliding time window (t_k, t_{k+s}) where $1 \leqslant k \leqslant l\text{-}s$ and obtain FCs defined on structurally-connected grey matter (SCGM) voxels. By concatenating all *FC*s into a vector, we thus generated the functional connectivity vector (*FCV*) of all the fibers defined at each time point *k*:

$$FCV(k) = FC\left(v_g, v_h, t_i, t_j\right) | (v_g, v_h) \in SCGM, t_i = k, t_j = k + s \quad (2)$$

FCV consists of *m* (total number of fiber-connected voxel pairs) elements, each is the connectivity strength of a specific voxel pair within the time window. Therefore, for each brain, we can extract (*l*-*s*) *FCV*s. One running example is given in Fig. 3.

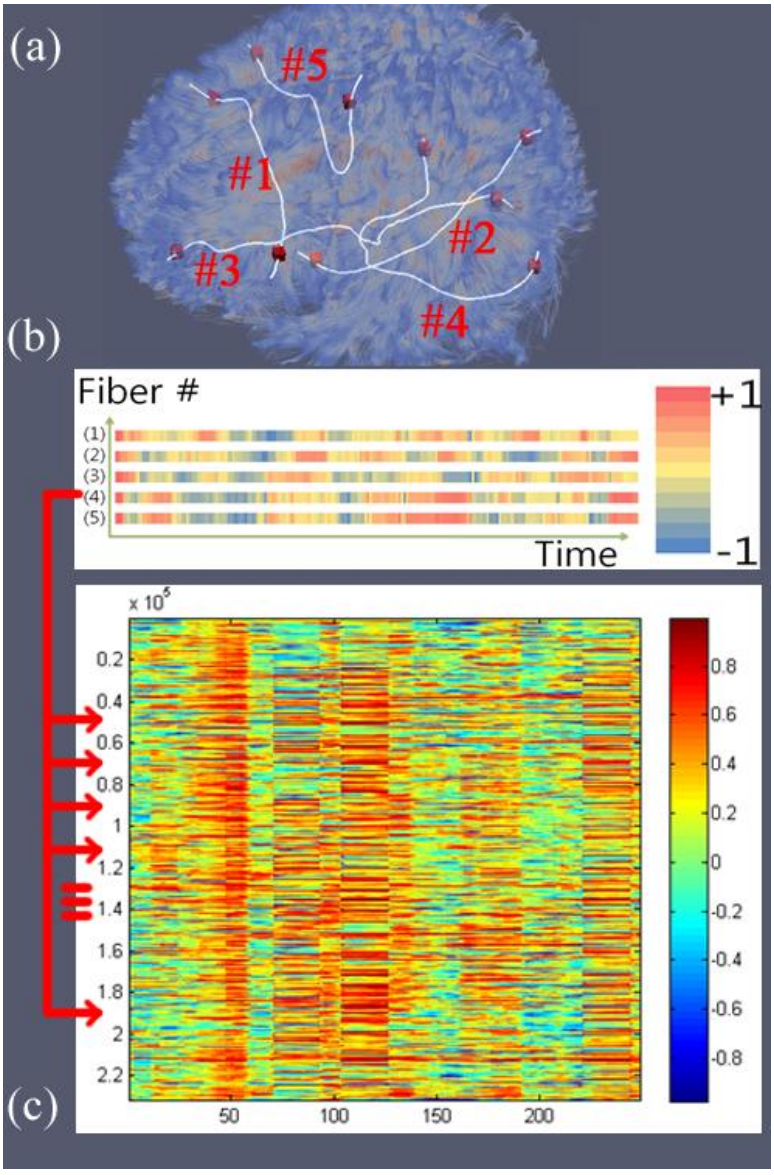

Fig.3. (a) Fibers with their end points in cortical gray matter, five fiber connections are highlighted; (b) Dynamics of functional connectivity of the above five SCGM voxel pairs. The temporal correlation between a specific voxel pair within each time window is

a single cell in the corresponding color-coded vector (1 to 5). Thus each color-coded vector is a visualization of the connectivity strength dynamics of that voxel-pair; (c) Combined *FCV*s through the whole time course, which is an extension of (b) from five SCGM voxel pairs to all SCGM voxel pairs

The *FCV* model has the capability to characterize and describe the dynamics of functional brain states based on multimodal DTI/fMRI data, and have been applied on task-based fMRI [37], resting state fMRI [38], and natural stimulus fMRI data sets [39]. Meaningful and promising results were obtained from the experiments. In particular, our results have shown that the functional brain state change curve roughly follows the external stimulus paradigm used in task-based fMRI as shown in Fig. 4, which partially validates our approach in that our algorithmic pipeline is totally data-driven and no a priori knowledge was used in the analysis.

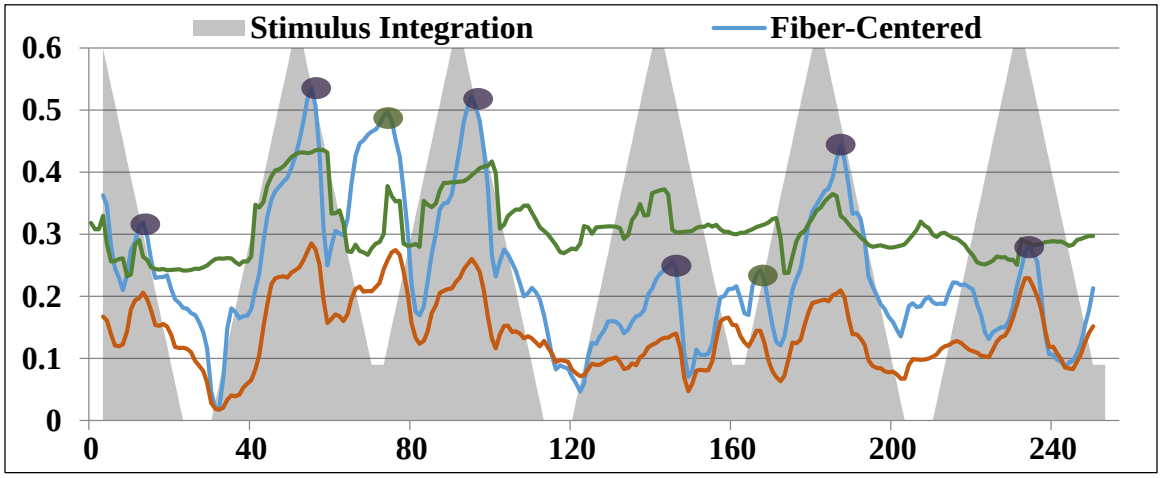

Fig.4. Temporal alignment between the stimulus curve and the global functional connectivity dynamics. The horizontal axis represents the temporal points of brain activation; vertical axis is the averaged functional correlation value (except for integrated stimulus function). Global functional connectivity for fiber-connected voxels is shown as

blue curve, global functional connectivity from randomly connected voxels is shown as red curve, ROI-based inferred functional connectivity is shown as orange curve, and the integrated stimulus curve are shown as grey triangles.

The major contributions of the structurally-guided dynamic functional connectivity modeling above are three-folded. Firstly, we developed, validated and applied a novel fiber-centered approach to defining the functional connectivity pattern in the human brain, and proposed the *FCV* pattern to represent a functional brain state. Secondly, instead of using raw fMRI BOLD signals, we use the *FCV* pattern that measures and represents the whole-brain functional connectivity of fibers for brain state change detection. Thirdly, the work in this paper provides novel understanding of and perspective on the dynamic behaviors of functional brain connectivity, which cannot be seen in traditional static connectivity analysis, and offers a starting point for in-depth elucidation of the complex patterns of large-scale functional brain interactions.

However, due to the limitations in white matter fiber registration techniques, we have not obtained inter-subject correspondence on the *FCV* patterns in this model. As the result, the main objective of the model is for characterization, rather than detailed investigation or used for classification purposes. Such limitation is solved by the DICCCOL system we have later adopted, which will be introduced in section 2.2 below.

2.2 Temporal/Spatial Segmentation of fMRI Data

In order to provide the much-needed inter-subject correspondence, our group developed and validated 358 consistent and corresponding DTI-derived landmarks across

multiple brains and populations, named Dense and Individualized Common Connectivity-based Cortical Landmarks (DICCCOL) ROIs [14, 35]. In short, each ROI in the DICCCOL system was optimized to possess maximal group-wise consistency of DTI-derived fiber shape patterns. The neuroscience basis is that each brain's cytoarchitectonic area possesses a unique set of intrinsic inputs and outputs, namely the "connectional fingerprint" [40], which largely determines the functions that each brain area performs. This close relationship between consistent structural connection pattern and brain function has been replicated in a series of our recent studies [41]. This set of 358 DICCCOL landmarks has been reproduced in over 240 brains of four separate healthy populations. Importantly, this set of 358 DICCCOL landmarks can be accurately predicted in an individual subject based only on DTI data. In this way, the voxel-wise fMRI data in each subject is transformed into a set of signals defining on 358 ROIs. Further, inspired by the success of Bayesian graphical causal models in neuroimaging (e.g., [42-44]) and in recognition of the importance of revealing dynamics of functional interaction patterns as introduced in section 2.1, we then developed the dynamic Bayesian variable partition model (DBVPM) that simultaneously infers global functional interactions within brain networks and their temporal transition boundaries. A key conceptual novelty in DBVPM is that the temporal boundaries of functional brain activities represented by fMRI time series are defined and determined by the abrupt change points of multivariate dependences among networks, instead of the fMRI time series changes or pair-wise functional connectivity changes as used in section 2.1. Thus the temporal stationarity of functional interaction patterns within each time segment is achieved automatically and the temporal boundaries of successively different multivariate functional interaction patterns are statistically inferred naturally

within a unified Bayesian framework. The overall two-level MCMC framework is illustrated in Fig. 5.

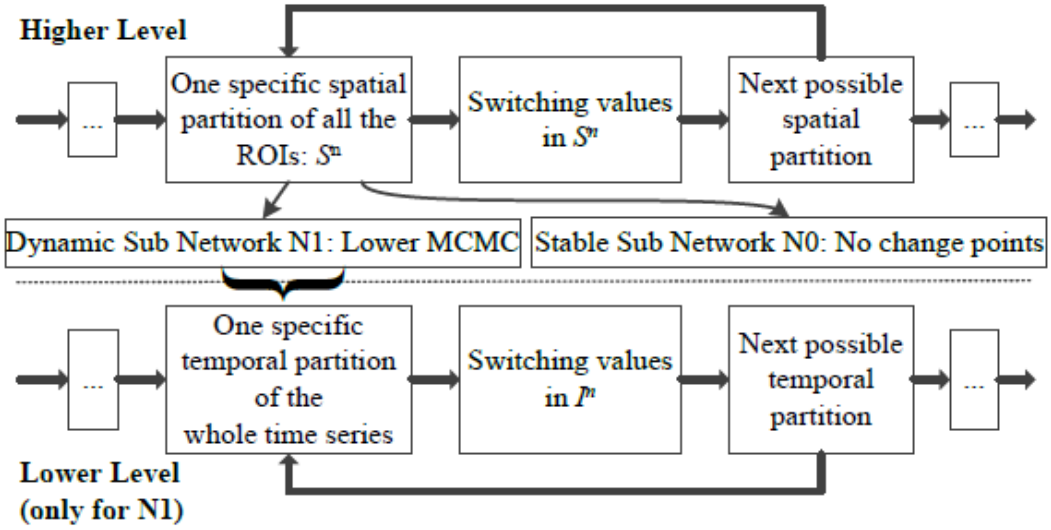

Fig.5. Algorithmic pipeline of the two-level MCMC scheme for simultaneous functional interaction and transition pattern inference.

Specifically, the Bayesian variable partition model aims at estimating the posterior distribution of the independent and identically distributed observations from the *d*-dimensional multivariate normal distribution [y_1, y_2, …, y_m]. By calculating the marginal distribution of the data [y_1, y_2, …, y_m]:

$$P(y_1, y_2, ..., y_m) = \frac{P(y_1, y_2, ..., y_m; \mu, \Sigma)}{P(\mu, \Sigma \mid y_1, y_2, ..., y_m)} = (\frac{1}{2\pi})^{md/2} \left(\frac{\kappa_0}{\kappa_m}\right)^{d/2} \frac{\Gamma_d\left(\frac{\nu_m}{2}\right)}{\Gamma_d\left(\frac{\nu_0}{2}\right)} \frac{(\det(\Lambda_0))^{\nu_0/2}}{(\det(\Lambda_m))^{\nu_m/2}} 2^{md/2} \quad (3)$$

Given the marginal distribution, we then defined the chain-dependence model if the variables can be partitioned into three non-overlapping sub-groups A, B, and C such that A and C are conditionally independent given B. Similarly, we defined the V-dependence

model if A and B are marginally independent, while C can be viewed as "children" of A and B. The definition of the two models are given in Eq. 4:

$$
\begin{aligned}
P(A,B,C) &= P(A)P(B\mid A)P(C\mid B) \\
&= P(A)\frac{P(A,B)}{P(A)}\frac{P(B,C)}{P(B)} \\
&= \frac{P(A,B)P(B,C)}{P(B)} \\
&= \frac{F(A,B)F(B,C)}{F(B)}
\end{aligned}
\qquad
\begin{aligned}
P(A,B,C) &= P(A)P(B)P(C\mid A,B) \\
&= P(A)P(B)\frac{P(A,B,C)}{P(A,B)} \\
&= \frac{F(A)F(B)F(A,B,C)}{F(A,B)}
\end{aligned}
\qquad (4)
$$

Given the dependence structure as chain (S=1) or V (S=0) structure, we can calculate the posterior distribution correspondingly. Then we designed the Markov chain Monte Carlo (MCMC) (Me-tropolis-Hastings) to sample from the posterior distribution with the following proposals (essentially searching for the optimized solution):

1)	Randomly choose one ROI and change its subgroup membership;
2)	Randomly choose two ROIs and switch their subgroup memberships;
3)	Swap the values of S (either from 0 to 1, or from 1 to 0).

The MCMC above constitutes the spatial partitioning model. In a similar way we can define the temporal partitioning model to investigate the dependency structures among the ROIs between different time periods and where are the boundaries of temporal blocks that exhibit significant differences from each other. Once these boundaries are determined statistically, they are considered as the change points of functional interaction patterns within the brain networks, as shown in Fig. 6

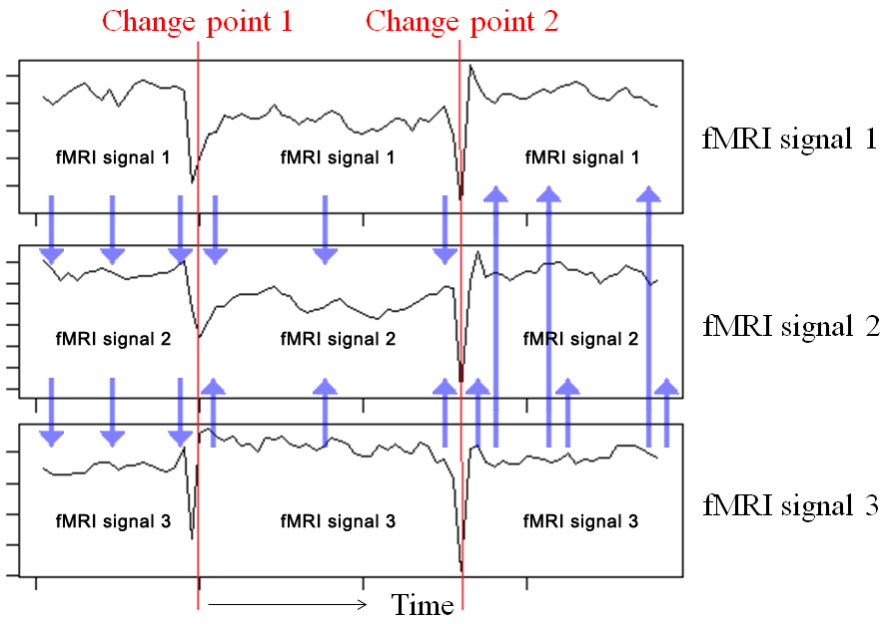

Fig.6. Illustration of two temporal change points of functional interaction patterns at time point 1 and 2. In the time period before change point 1, the functional interaction among three fMRI signals is a Chain-dependence model (signal1→signal2→signal3), while between time points 1 and 2, it is a V-dependence model (signal1→signal2←signal3). After the time point 2, it is again a V-dependence model (signal2→signal1←signal3).

In this way the DBVPM simultaneously models and characterizes high-order functional interactions and their temporal dynamics via a unified Bayesian framework. To solve the model, we then applied a two-level Metropolis-Hastings (MCMC) scheme to sample from the posterior distribution of the block boundaries and dependency structures within each block. The lower level MCMC samples from the posterior distribution of the dependency structures of each block given the block boundaries as discussed above, and the higher level MCMC samples from the posterior distribution of the block boundaries. Specifically, in the lower level MCMC, the proposal scheme involves alternating between the chain and

V structures and changing the group labels of each variable. In the higher level MCMC, the proposal scheme involves segmenting one block into two, merging two neighboring blocks and shifting a block boundary. In each higher level step, every block runs through a lower level MCMC. A dependency structure is sampled for each block as the dependency structure for that block in the higher level proposal. Then the log likelihood of the proposal can be calculated by summing up the log likelihood of each block. We then check the mixing of MCMC and exclude the burn-in from the actual MCMC sample of the posterior distribution. Then the posterior probability for each time point to be a change point was calculated from MCMC samples. A running example of how the DBVPM reveals the dynamics of functional interaction patterns is shown in Fig. 7.

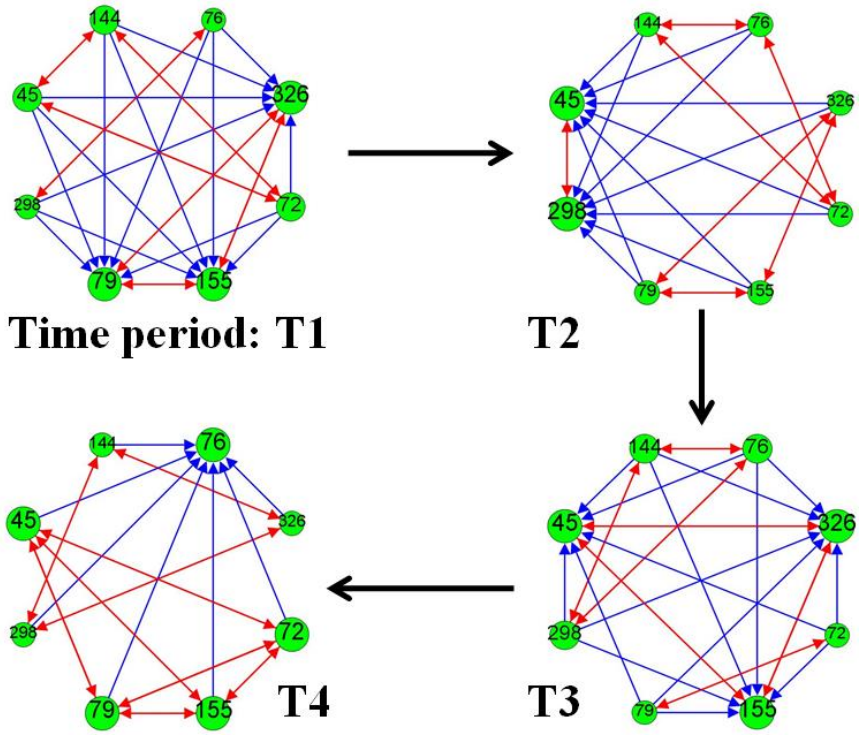

Fig.7. A running example of the evolving dynamic functional interaction patterns within the DMN. Blue lines: directed edges; red lines: undirected edges; green spheres: brain regions represented by DICCCOL.

The DBVPM was evaluated and validated with simulated data and obtained very accurate temporal/spatial segmentation of the data. One example result is shown in Fig. 8.

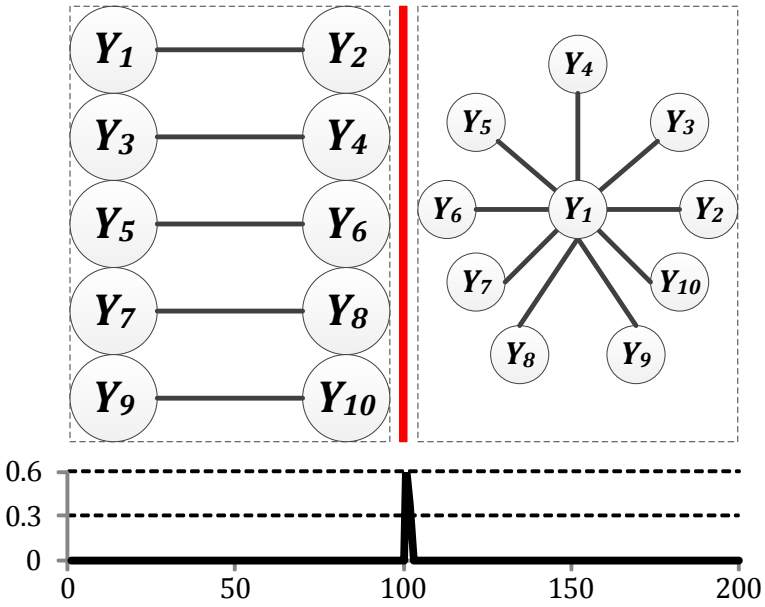

Fig.8. Results on the simulation data. Each node denotes one ROI, with the change of their contently structure at time point 100. Detected change point is shown below.

For the purpose of testing its capability on real fMRI data, we used the functionally labeled DICCCOLs and their prediction models to locate the default mode network (DMN) ([45-47]) and emotion network [48] from a post-traumatic stress disorder (PTSD) dataset. The results revealed substantially different multivariate functional interaction signatures and temporal transitions in the default mode and emotion networks of PTSD patients, in comparison with those in healthy controls. This result demonstrated the effectiveness and utility of DBVPM in elucidating interesting features that can-not be revealed by traditional static pair-wise functional connectivity analysis. In summary, the study shows that Bayesian graph is a powerful tool for modeling connectivity and its dynamics, and the

change point detected by the model provides the statistical foundation for our further dynamic functional network analysis.

2.3 Functional Networks Dynamics as Connectomics

Based on the studies in section 2.1 and 2.2, we have successfully characterized and modeled the functional brain dynamics and found the method to automatically identify the change points for brain functional states. However, there is still a lack of integrated framework to describe the connectivity level at group-level across multiple subjects. As it has been recognized in neuroimaging literatures that functional connectomes constructed via neuroimaging data could offer a complete description of the macro-scale structural connectivity within the brain (e.g., [30, 33, 49-51]). In addition, functional connectomics signature have been shown to be powerful in characterizing and differentiating brain conditions [20]. However, most of the such works applied the assumption of functional stationarity which is contrary to the dynamics nature of the functional brain, as discussed in section 2.1. Therefore, quantitative modeling and characterization of functional brain dynamics has been of general interest in the neuroimaging community for years. For instance, functional microstates (a concept similar to the functional brain states in our fMRI study) have been well-established in EEG data modeling and analysis [52-55]. Thus it is generally agreed that quantitative characterization of these time-dependent functional connectivity/connectome dynamics and representative patterns can elucidate fundamentally important temporal attributes of functional connections that cannot be seen by traditional static pairwise functional connectivity analysis [16].

From a technical perspective, the discovery, cross-validation and application of EEG-based microstates have been fundamentally enabled and facilitated by a standardized electrode reference system such as the International 10/20 System. As a result, the measured EEG signals and the identified dynamic microstates in different brains can be readily mapped into a standard reference system, and thereby effectively integrated and compared. However, for R-fMRI data, it has been challenging to integrate and compared fMRI signals and their derived measurements across different brains due to the lack of a reliable and accurate brain localization and reference system. Thus in our work, similar to the approach in section 2.2, the DICCCOL [14] system is adopted for predicting the ROIs from the individual brains. As the DICCCOL ROIs possess intrinsically-established structural and functional correspondences across individuals, they provide a natural general brain reference system across individuals and populations. The structural and functional connectomes established by DICCCOL is then used to define and characterize functional microstates based on R-fMRI data. In particular, we used the large-scale functional connectivity among DICCCOLs as functional connectome (*FC*), and manually divided the temporally varying *FC*s into quasi-stable segments. For instance, a typical R-fMRI scan with time length of 10 minutes can be segmented into 10-20 quasi-stable *FC* segments, within which the *FC*s are averaged into one vector. The whole analytics pipeline is illustrated in Fig. 9.

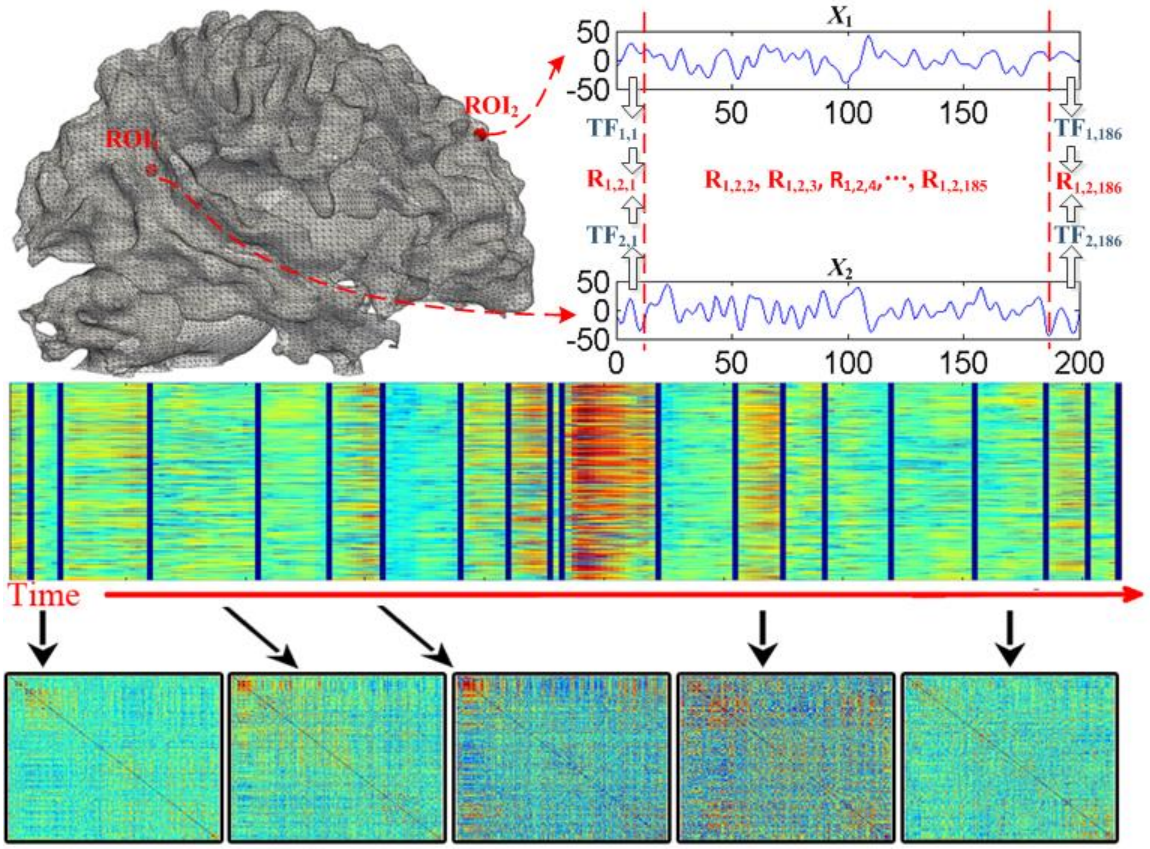

Fig.9. Illustration of calculating dynamic *FC* and *FCS*. Top panel: Illustration of R-fMRI signals extracted from two DICCCOL ROIs of a randomly selected subject, and their sliding time-windowed functional correlations. Middle panel: Functional connectome strength (*FCS*). Each column is a single FCS_t at one window, and each row is the dynamics of the connectivity strength of one DICCCOL ROI. Bottom panel: Averaged FCs of segmented states as color-coded matrices at the corresponding brain states.

Specifically, we applied a sliding time window approach to divide the extracted R-fMRI signal X_i from the *i*-th DICCCOL into temporal segments $TF_{i,t}$ at time point *t*, with the duration of window length *l*:

$$TF_{i,t} = X_{i,p} | t \leq p < t + l \tag{5}$$

Then for every pair of temporal segments $TF_{i,t}$ and $TF_{j,t}$ of the time series X_i and X_j from two DICCCOLs, we calculated the Pearson correlation $R_{i,j,t}$ between them:

$$R_{i,j,t} = corr\left(TF_{i,t}, TF_{j,t}\right); R_{i,j,t} = 0, \text{if } i = j; FC_t = R_{i,j,t} | i, j \in (1, 358) \quad (6)$$

where FC_t is the functional connectome (*FC*) at time *t*, which is a set of correlations from the congregation of all $R_{i,j,t}$ over every combination of *i* and *j*, to characterize the whole-brain functional connectivity. It is a symmetric matrix of dimension 358×358. For dimension reduction, we defined the functional connectome strength (*FCS*):

$$FCS_t = \sum_{j=1}^{358} R_{i,j,t}\,;\; FCS = \{FCS_1, FCS_2, \ldots, FCS_{T-l}\} \quad (7)$$

where FCS_t is the summation of correlations of each DICCCOL ROI with all the other ROIs at time *t*, which is an *I*×1 vector. Thus the *i*-th value in the vector is the strength of connectivity of the *i*-th ROI. *FCS* is the aggregation of FCS_t, representing the dynamics of connectome strength through the entire time course. Given that the averaged *FC*s possess intrinsically-established correspondences across individuals, all of the averaged FCs from different individuals are then pooled and clustered into Common Functional Connectomes (CFCs) via the Fisher Discriminative Dictionary Learning (FDDL) algorithm [56]. The basic idea of FDDL is to learn a structured dictionary *D* from the training data *A* so that *A*=*DX*, where *X* is the coding coefficient. *D* contains certain numbers of sub-dictionary D_i that corresponds to the class labels in the training data, under the constraint of maximizing the discriminative capacity of the dictionary. Then, the following energy function *J*(*D*,*X*) were optimized to obtain the learned dictionary *D* and its corresponding projection *X* of the data on *D* [56]:

$$J_{(D,X)} = argmin_{(D,X)}\, r(A, D, X) + \lambda_1 \|X\|_1 + \lambda_2 f(X); \quad (8)$$

$$r(A, D, X) = \|A_m - DX_m\|_F^2 + \|A_m - D_m X_m^m\|_F^2 + \sum_{\bar{m}}^{c} \|D_{\bar{m}} X_{\bar{m}}^m\|_F^2$$

$$f(X) = tr(\sum_{m=1}^{c} \sum_{x_{k \in X_m}} (x_k - \overline{X_m})(x_k - \overline{X_m})^T - tr(\sum_{m=1}^{c} n_m (\overline{x_m} - \overline{X})(\overline{x_m} - \overline{X})^T) + \eta \|X\|_F^2$$

The first term in the energy function, $r(A, D, X)$ is the constraint on discriminative fidelity, allowing the dictionary D able to code the data A (which is the congregated *FCS* matrix) with minimum residual, while at the same time only using one sub-dictionary D_m, but not other sub-dictionaries. The neuroscience rationale behind it is that each sub-dictionary D_m learned is corresponded to one way of the classification of *FCS*, where within this class the *FCS* are similar with each other, and X is the projection of the dataset A on dictionary D, thus it is the classification result (class labels). In Eq. 7, X_m is the projection of A_m, which is one of the sub-classes in A, on the whole dictionary D. X_{mm} is the projection of A_m on the correct sub-dictionary D_m. Thus to minimize Eq. 8, we will optimize dictionary and its corresponding classification to use the correct sub-dictionary D_m to project A_m, and avoid using other sub-dictionaries. The second term in the energy function in Eq. 8 is the sparse constraint, requiring the coding coefficient X be as sparse as possible, i.e., the total number of non-zero items in X should be minimized. With this constraint, each single *FCS* in A would only be projected by a limited number of sub-dictionaries, which is in accordance with our premise that the brain states would be discrete with abrupt change on the boundaries. The third term $f(X)$ is the constraint on the discriminative coefficient, which aims to minimize within-class scatter of X, and maximize cross-class scatter of X, according to Fisher discrimination criterion. In the definition of $f(X)$, X_k is the item (single *FCS*) in X_m i.e. each class of the projection. The within-class scatter of the projection was measured

by the summed distance from each item in the class (x_k) to the average of the class. The cross-class scatter was measured by the summed distance from the average of each class to the average of the whole data. The integer c is the total number of classes in the training data A, as well as the total number of sub-dictionaries in dictionary D. The integer n_i is the number of items in X_i, as there are multiple items in each class of the training data A, and η is the scaling constant. The learned dictionary has the same dimension with the input training matrix A, and contains sub-dictionaries corresponding to each class label. Also, as every state has its own *FC*, as depicted in Figure 8, we obtained the average *FC* for each sub-dictionary, defined as *CFC* in our study. The dynamics of brain connectomes could then be projected into the small number of representative *CFC*s with minimum information loss. The learned dictionary was then applied to classify the testing matrix T, to project brain state in testing data to the *CFC*s. The solution for the above sparse coding algorithm is based on the sparseness function developed in [57] and further enhanced in [58], which are all well-established methods.

With the hypothesis that functional connectomics signatures can effective characterize and differentiate healthy control and mental disorders, we applied the model on the PTSD dataset as used in section 2.2. Experimental results showed that the *CFC*s patterns are remarkably reproducible across healthy controls and PTSD patients, as visualized in Fig. 10 below.

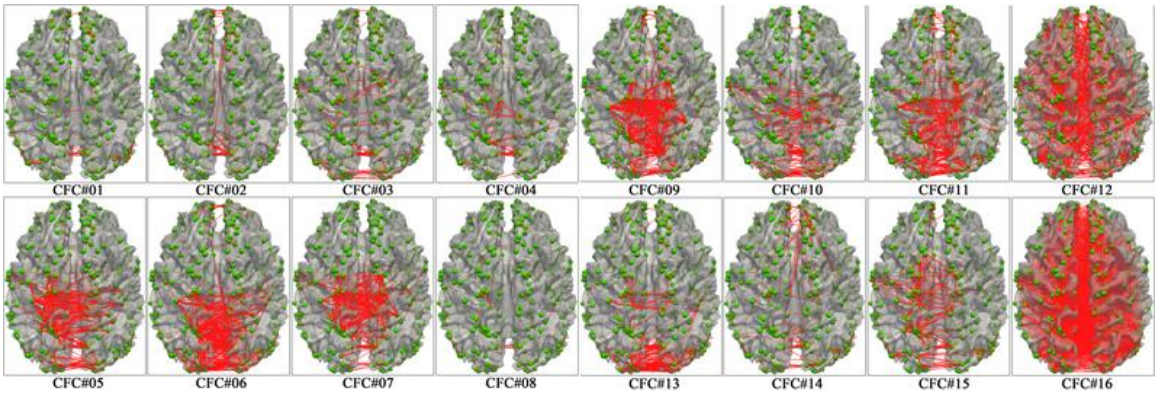

Fig.10. Visualization of *CFC*s on cortical surfaces. DICCCOL ROIs are marked as green nodes on the cortical surface. Functional connectivities between ROIs are shown as red edges connecting those nodes.

More interestingly, two additional *CFC* patterns with altered connectivity patterns (termed signature functional connectome (*SFC*)) exist dominantly in PTSD subjects. More importantly, these two *SFC* patterns alone can successfully differentiate 80% of PTSD subjects from healthy controls with only 2% false positive. These results suggest that *SFC* patterns could be potentially used as the biomarkers of PTSD in the future. Furthermore, based on the clustered CFCs, the time series R-fMRI data was projected into a series of temporally concatenated CFCs and their temporal transitions patterns were modeled by a finite state machine (FSM). Essentially, the graph structure and the edge connection strength of the FSM characterize the probability of transition from one CFC pattern to another. Our experimental results revealed that meaningful and reproducible FSMs can be learned from separate groups of brains. In particular, it was found that the FSM learned from PTSD subjects exhibits a substantially altered pattern of CFCs and transition patterns, in comparison with the healthy controls, as illustrated in Fig. 11. These results suggest that not only the CFCs patterns themselves, but also their temporal transition patterns, can

contribute to the characterization and differentiation of PTSD subjects from healthy controls.

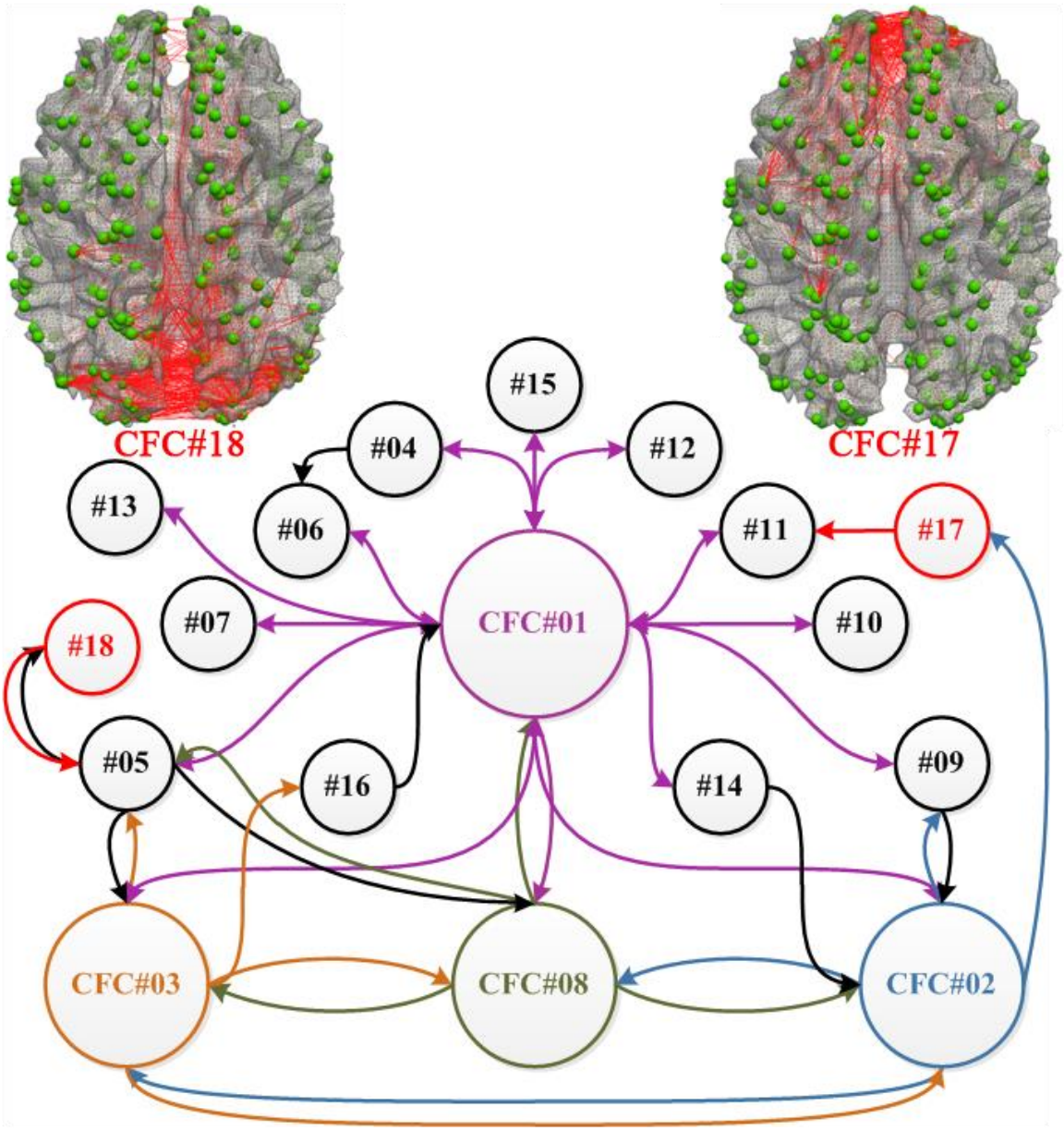

Fig.11. The state flow diagram depicting significant transitions between *CFC*s of PTSD patient subjects. Nodes representing state #17 and #18 are colored in red as they are *SFC*s for PTSD. The patterns of functional connectivities of CFC #17 and #18 on cortical surface are shown to the top panel.

In summary, the common functional connectome (*CFC*) and the derived signature functional connectome (*SFC*) modeling has been shown to be capable of summarizing the functional connectivity from multiple time fragments of multiple subjects. The result effectively encodes the functional brain dynamics into the label of *CFC*s and can be used to differentiate subjects from different groups, especially between healthy and patients. Using the similar approach, we have also identified the shared and signature connectomics for the functional brain between during resting-state and during task [2].

CHAPTER 3

FUNCTIONAL NETWORK DISCOVERY

BY DICTIONARY LEARNING METHOD

3.1 Component-based Analysis for Functional Network Study

As described in Chapter 2, we have characterized and modeled functional connectivity and the group-wise connectomics through various approaches including pair-wise connectivity measurement, spatial/temporal segmentation and group-wise connectomics representation. The next question for better understanding the organizational architecture of cortical function is whether we can infer the similar dynamics pattern directly from the voxel-wise fMRI data, rather than relying on the connectivity measurements. Such question is especially important as there has been mounting evidence [59-63] that the brain function emerges from and is realized by the interaction of multiple concurrent neural processes or networks, each of which is spatially distributed across specific structural substrate of neuroanatomical areas [64]. However, it is still challenging to robustly and faithfully reconstruct concurrent functional networks from raw voxel-wise fMRI data and quantitatively measure their network-level interactions, mainly due to the huge data size and high noise level, as comparing with the more concise connectivity measurements used in Chapter 2. Traditionally, the subtraction approach (contrast between task and baseline epochs) has been the dominant methodology in tfMRI paradigm design and tfMRI data analysis [65], based on which a majority of previous human neuroimaging/brain mapping studies and conclusions were derived and rooted. Despite the remarkable successes and

significant neuroscientific insights achieved by the subtraction approach, nevertheless, it has difficulty in reconstructing concurrent, interacting functional networks such as task-evoked and resting state networks. It has been recognized and pointed out in the literature that spatially overlapping networks sub-serving different functions are possible to go unnoticed by the blocked subtraction paradigms and the associated analysis methods such as general linear model (GLM) [66, 67]. Responding to the limitations in the activation detection-based methods, as well as for the application on resting-state fMRI data, recently a variety of computational methods such as independent component analysis (ICA) [68], normalized cut [69] or other clustering algorithms [70] have been employed to map functional networks in healthy brains or neurological/psychiatric disorders. Among these methods, the component-based modeling has been widely applied in recent studies due to its versatility and high reproducibility for group-wise inference [15, 16, 71, 72]. In the component-based modeling, the fMRI signal matrix is used as an input for the matrix decomposition with specific regularizations (e.g. independence), while the decomposing results (i.e. "components") are then interpreted as functional networks.

3.2 Sparsity-regularized Matrix Decomposition (Dictionary Learning)

It should be pointed out, however, that current functional network identification methods did employ the strategy of spatially clustering fMRI signals [68, 73]. Such assumption is that RSNs are not spatially overlapping with each other, that is, the networks are spatially independent as much as possible. In order to address the abovementioned questions and bridge the current significant neuroscience knowledge gaps, we developed a novel computational framework of identifying functional networks from the whole-brain

voxel-wise fMRI signals by sparsity-regularized matrix decomposition method (i.e. dictionary learning). Inspired by the successes of using sparse representation for signal and pattern analysis in the machine learning and pattern recognition fields [74], the basic idea of the framework is to aggregate all of hundreds of thousands of fMRI signals within the whole brain of one subject into a big data matrix, which is subsequently factorized into an over-complete dictionary basis matrix and a reference weight matrix via an effective online dictionary learning algorithm [75]. Specifically, for the input matrix $S \epsilon R^{t \times n}$, where t is the total number of time points and n is the number of voxels, we aim to learn a meaningful and over-complete dictionary $D \epsilon R^{t \times m}$ ($m>t$, $m<<n$) [75] for the sparse representation of S based on the cost function l:

$$\ell(S, D) \triangleq \min_{D \epsilon C, \alpha \epsilon \mathbb{R}^{m \times n}} \frac{1}{2} ||S - D\alpha||_F^2 + \lambda ||\alpha||_1 \quad (9)$$

In this way, the input matrix from a subject's whole-brain voxel-wise fMRI signal will be represented by a learned dictionary matrix and a sparse coefficient matrix, as illustrated in Fig. 11. Each column of the α matrix contains the sparse weights for interpreting each fMRI signal with the atomic basis signals in the dictionary. Meanwhile, each row of the α matrix stores the information of the voxel spatial distributions that have references to the corresponding dictionary atoms. The time series of each over-complete basis dictionary represents the functional BOLD activities of a brain network. A particularly important characteristic of this framework is that the decomposed reference weight matrix naturally reveals the spatial overlap/interaction patterns among reconstructed brain networks. Extensive experimental results demonstrated that this novel methodology can effectively and robustly uncover multiple functional networks, including both task-evoked and RSNs, which can be well-characterized and interpreted in spatial and/or temporal domains.

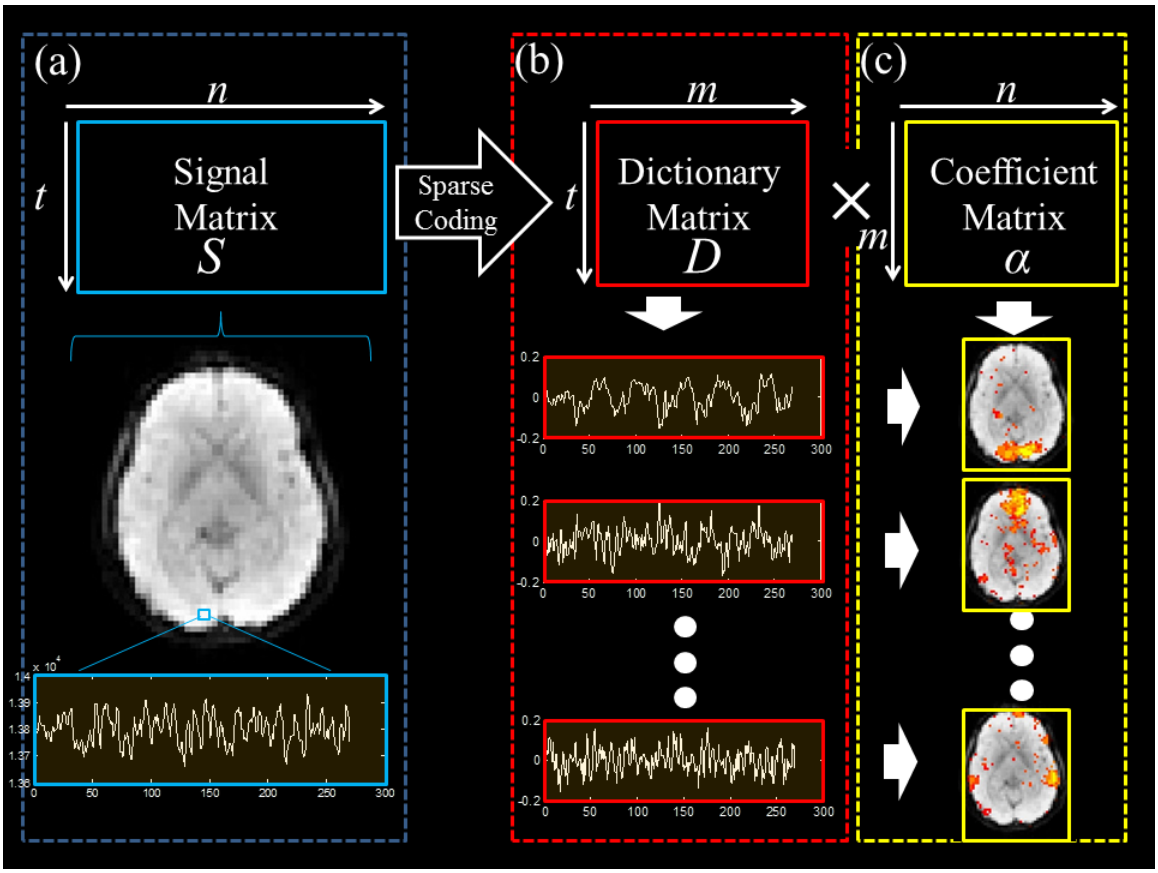

Fig.12. The computational pipeline of sparse representation of whole-brain fMRI signals using dictionary learning. (a) The whole-brain fMRI signals are aggregated into the data matrix. (b) Illustration of the learned atomic dictionary. Time series of three exemplar components are shown in the bottom panels. (c) The decomposed reference weight matrices, each row measures the weight parameter of each component in the whole brain.

With the decomposed dictionary components and their reference weight parameters across the whole brain for each subject, our next major task is to characterize and interpret them within a neuroscience context. In our study, we applied the dictionary learning method on publicly released large-scale Human Connectome Project (HCP) high-quality tfMRI data (Q1 release) [17]. Experimental results from the HCP datasets shown that these

well-characterized functional networks are reproducible across different tasks and individual brains and exhibit substantial spatial overlaps with each other, thus forming a collection of holistic atlases of functional networks and interactions (HAFNI). In our experiments, functional networks from different individuals are manually aligned by the cross-examination of several experts, one example is shown in Fig. 13 and the group-wise results are provided in Fig. 14.

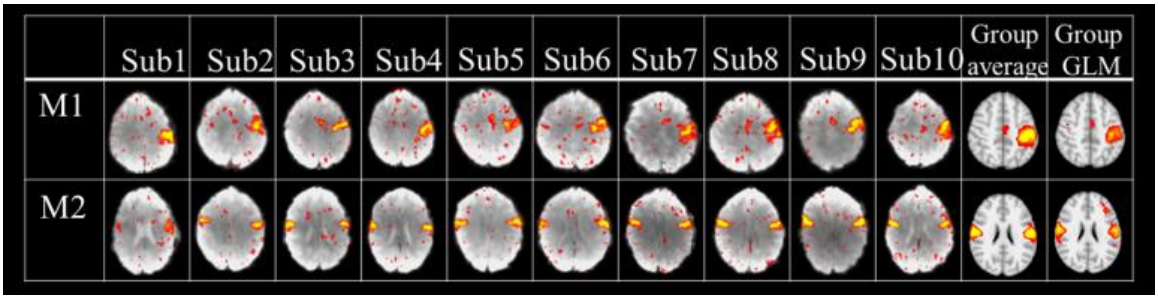

Fig.13. The task-evoked HAFNI components and the comparison with GLM-derived activation maps. HAFNI components across 10 different HCP subjects in the motor task are visualized in the two rows. The last two columns are the group-wise averages of HAFNI components and the group-wise GLM activation maps.

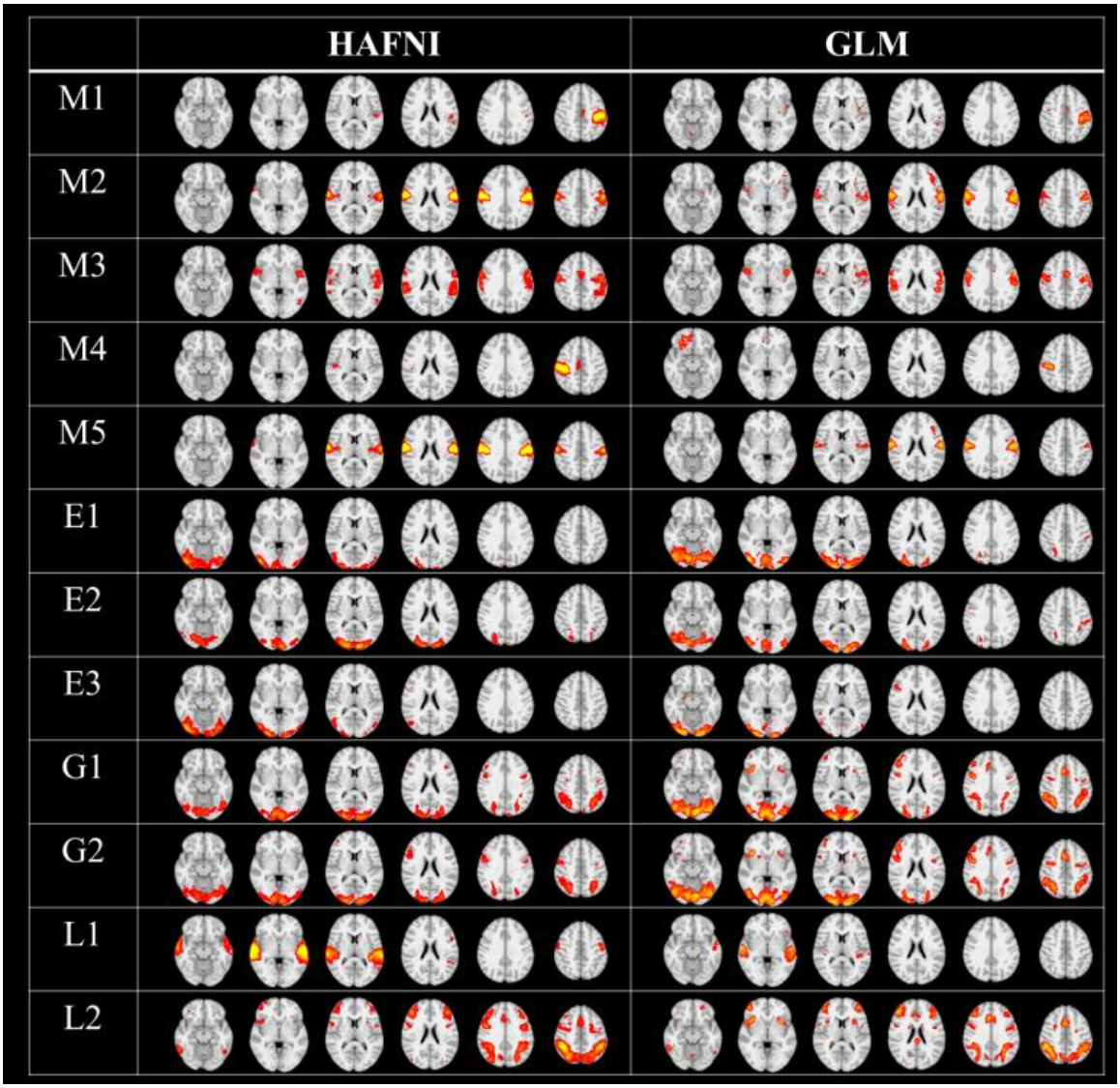

Fig.14. Group-wise averages of 12 HAFNI components across HCP subjects for the four tasks, including motor (M), emotion (E), gambling (G), language (L), as well as the corresponding averaged GLM-derived activation maps (rightmost column).

More interestingly, these HAFNIs revealed two distinct patterns of highly heterogeneous (highly overlapped) regions and highly-specialized (task-evoked) regions in tfMRI data and showed that these two patterns of areas are reciprocally localized. In general, our work suggests a novel organizational principle of human brain function: its

functional architecture is reciprocally composed of hybrid highly-specialized areas and highly heterogeneous areas. The reciprocal spatial distribution of the networks is illustrated in Fig. 15, showing the existence of both functional Highly-Heterogeneous Region (HHR) and Highly-Specialized Regions (HSR). HHR region can be considered as the multiple-demand (MD) area of the brain [59, 61], while an F-HSR region can be considered as a demand-specific (DS) area [59]. It is interesting that those two types of regions are also reciprocally distributed and widespread across the cerebral cortex, suggesting that the functional cortical architecture is composed of a reciprocal combination of frequent highly-specialized regions and frequent highly-heterogeneous region across different types of cognitive or functional tasks.

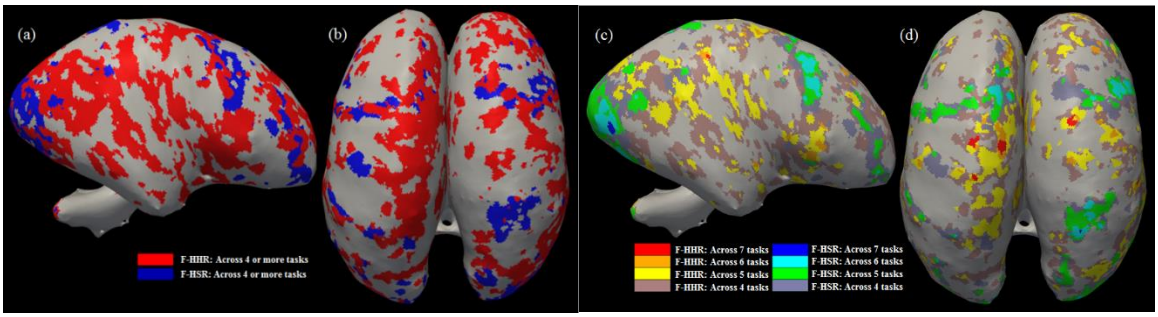

Fig.15. Visualization of the frequent HHRs (F-HHR) and frequent HSRs (F-HSR) for one individual brain across 7 tasks on the inflated cortical surface, color-coded according to the color bar at the bottom.

3.3 Functional Network Transition Modeling

Based on the functional networks discovered by the dictionary learning methods as introduced above, we can then model the transition among network dynamics through a sliding time-window based correlation-enforced learning scheme. Specifically, the input time series S (of dimension $T \times P$, T is the number of time points, P is the number of voxels)

is divided into *N* number of overlapping temporal segments (i.e. time windows), each with window length of *W*. The overlapping section between the two consecutive windows contains *W*/2 time points. For the first temporal segment S_1, the dictionary learning algorithm is employed to learn a dictionary D_1 and the corresponding weighting coefficient matrix z_1, which characterize the basis temporal variation patterns and the underlying sparse structures of S_1 as described in Eq. 9 using *K* number of dictionary atoms. For the later temporal segments S_n (n>1) we perform the similar dictionary learning analysis but with enforcement applied upon the learned dictionary D_n to maximize the similarity between the overlapping portion of D_n and D_{n-1}:

$$\ell(D_n, z_n) = \min_{D_n \epsilon \mathbb{R}^{W \times K}, z_n \epsilon \mathbb{R}^{K \times P}} \frac{1}{2} \|D_n z_n - S_t\|_F + \lambda_1 \| z_n \|_1 - \lambda_2 corr^*(D_n, D_{n-1}), \qquad (10)$$

where the *corr** function measures the atom-wise similarity between the overlapping section (denoted by D_1 and D_2) of the two dictionary matrix D_n and D_{n-1}. The final D_n is then the combination of D_1 and D_2. The gradient for *z* is unaffected by the enforcement term thus remains the same as in the original dictionary learning method. As the first *W*/2 number of time points in D_n and the last *W*/2 number of time points in D_{n-1} overlaps in time, considering a functional network with consistent activation patterns over time, the *corr** function for its corresponding atoms in dictionary D_n and D_{n-1} shall be very high. Thus by adding the negative correlation (balanced by λ_2) as the extra penalty term, we are aiming to identify the most temporally consistent atoms in D_n regarding to atoms in D_{n-1} and align them at the same index of the two dictionary matrices. The correlation enforcement scheme is motivated by the results from our previous work, where highly-correlated atoms were found from the consecutive-learned dictionaries. Based on Eq. 10, *K* number of functional networks series (termed as "network continuum") with consistent temporal patterns within

the overlapping section and individualized (variability allowed) spatial patterns can be learned from the segmented fMRI data.

Based on the dictionary learning results, the next step is to cluster the spatial patterns (as $P\times1$ vectors) networks continnums. The clustering algorithm is consisted of two layers for individual-wise and group-wise clustering. In the first layer, networks within each subject are clustered to obtain the individual-level representative cluster centers. As the result of the first-layer clustering, networks from each subject will be represented by a few cluster centers. In the second layer, cluster centers from all subjects are then pooled together and further clustered. The final cluster centers obtained by the second-layer clustering are the group-wise representative brain functional states. The cluster centers are then mapped back to each functional network based on the clustering results of both layers, encoding the functional networks into discrete indices as illustrated in Fig. 16.

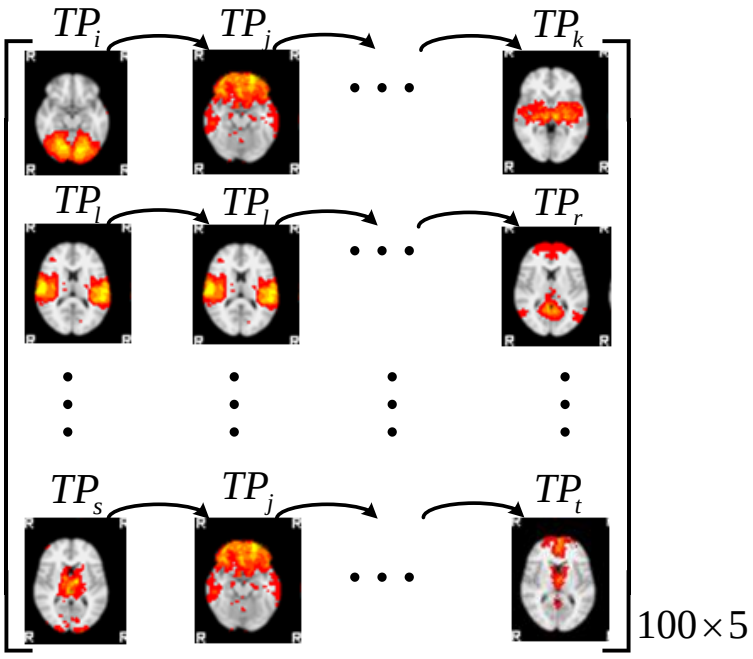

Fig.16. Encoded network transition matrix, entry is the representative functional state associated with the corresponding time points in the network continuum.

As the network dynamics of each subject are now encoded as the state matrix, the functional state transitions can then be characterized by the left to right changes between adjacent entries in the state matrix. Based on the state transition matrix, a weighted directed graph (termed as “transition graph”) is built, where significant transition events with higher frequency of occurrence are identified. The resulting significant transitions constitute the core transition graph. Using the core graphs obtained from healthy control and patient populations as features, we can then build the classifier to discriminate the two populations based on their functional state dynamics.

The experimental results by applying the model on the Autism Brain Imaging Data Exchange (ABIDE) datasets show that 74 discriminative transition types among 109 functional states can be extracted, which shows significant difference between the ASD patients and normal controls. 32 of them show significantly higher ($p<0.05$) frequency in ASD patients and 42 of them show significantly higher frequency in healthy controls, as summarized in Fig.16. While the transition graph of the two populations share the same set of nodes (i.e. functional states), their functional transitions are quite different which reveal the potential underlying functional dynamics changes as the result of altered brain functional organization pattern caused by the disorder.

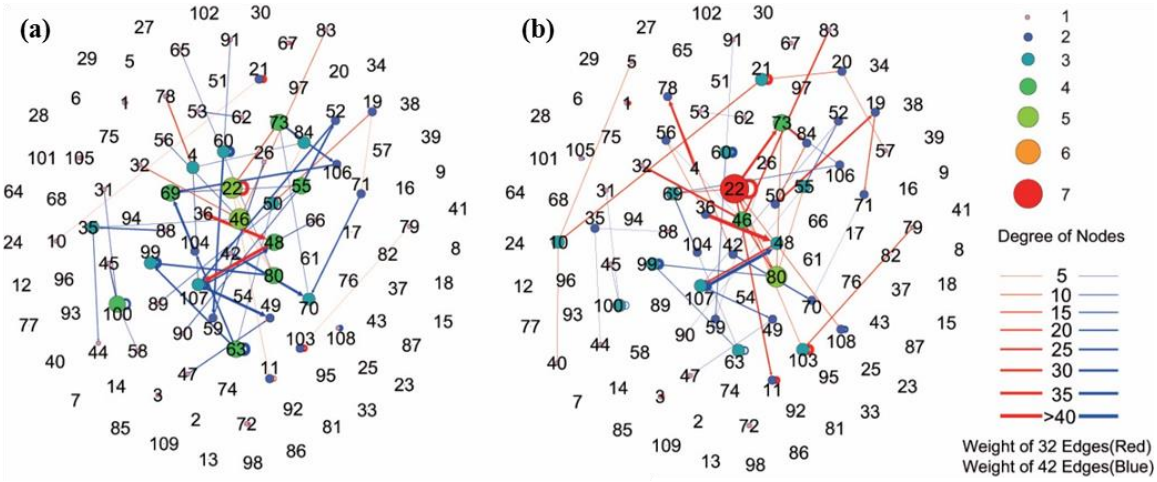

Fig.17. Visualization of the 74 discriminative functional transitions within the (a) normal control and (b) ASD patient populations. Each node in the graph represents one of the 109 functional state, color-coded by their frequency. Edges representing ASD-specific transitions are colored by red, healthy control-specific transitions are colored by blue.

Using the 74 discriminative transitions as feature vectors, the classifier is trained with 10-folds cross-validation for verification and validation. The classification performance is evaluated by average of the 10-folds results based the percentage of TP (true positive), TN (true negative), FP (false positive) and FN (false negative) as summarized in Table. 1. The classification performance shows that the proposed classification framework based on functional network transitions could successfully differentiate the healthy control and patient populations.

Accuracy	**ASD Patients as Positive Cases**		**Healthy Controls as Positive Cases**	
	Precision	**Recall**	**Precision**	**Recall**
94.0%	92.2%	97.0%	95.5%	93.0%

Table.1. Performance of the network transition-based classification framework.

3.4 Fast Dictionary Learning Method using Rank-1 Decomposition

As introduced in section 3.2, we have successfully applied dictionary learning method on the whole-brain voxel-wise fMRI data for functional network discovery. However, as the data size of fMRI quickly out grows the current computational power, current dictionary learning methods are still in need of more improvements regarding to their efficiency and scalability. Specifically, while the state-of-the-art methods such as online dictionary learning [75] and stochastic coordinate coding method [76] have been extensively explored in our prior studies, it has been found that these methods would take 10~20 minutes (depending on the hardware configuration) to finish the learning on one individual dataset consisting of around 1 million fMRI signals. As there are usually multiple subjects during multiple tasks/resting-state in each fMRI database (e.g. HCP Q1 release consisted of the fMRI data from 68 subjects during 7 different tasks and resting-state), it will then take tens of hours or even several days on our server machine to process the whole database, thus severely impeding the analysis progress. The bottleneck of the computational speed will be further worsened by another two facts. First, most of the previous dictionary learning works for fMRI data modeling assumed the temporal stationarity. That is, time series from the whole fMRI scan are used as the input for the learning without considering possible temporal dynamics of brain function. While it has been shown in various literatures that the organization of brain functions is under time-to-time state changes [3, 13]. In order to better capture the functional networks dynamics, it is more appropriate to perform the dictionary learning method on the temporal segment within each quasi-stable brain state. However, such approach will greatly increase the computational load as there would be fMRI data defined in multiple temporal segments to

be analyzed for each individual dataset. Second, individual-level or small group-level analysis might be insufficient for identifying the consistent and discriminative functional networks (e.g. for diagnosis purposes), limited by the cross-subject/cross-study variability in the reported results. In such cases, population-level studies taking the aggregated fMRI data from large number of subjects as input, e.g. the "1000 functional connectome" data [32] can overcome such limitations and provide more powerful inference for cognitive and clinical neuroscience studies. Yet as discussed in [77], the memory requirement of the population-level study quickly goes unbearable as database grows larger, thus requiring a more scalable and light-weighted computational framework.

To provide the solution targeting a more efficient and scalable dictionary learning method as discussed above, here we present the rank-1 dictionary learning (r1DL) model for the functional network decomposition from fMRI data. The model aims to iteratively estimate multiple rank-1 basis vector u (T×1 vector with unit length) and its loading coefficient vector v (P×1 vector) to decompose the input signal matrix S of dimension T×P by minimizing the following energy function $L(u, v)$:

$$L(u,v) = \|S - uv^T\|_F, \text{s.t.} \|u\| = 1, \|v\|_0 \leq r. \quad (10)$$

Eq. 10 indicates that the product of u and v is supposed to well-fit the input S while the total number of non-zero element in v should be smaller or equal to the given sparsity constraint parameter r. Note that Eq. 1 is a special formulation of the more general dictionary learning framework described in Eq. 9. The learned dictionary vector u describes one underlying functional activation pattern that serves as the basis for the whole brain functional dynamics, while the loading coefficient vector v describes how the basis u contributes to the activation at each voxel. The minimization problem in Eq. 10 can be

solved by alternatively updating u (randomly initialized before the first iteration) and v until convergence:

$$v = \underset{v}{\operatorname{argmin}} \|S - uv^T\|_F = (u^T S)^T, s.t. \|v\|_0 \leq r, u = \underset{u}{\operatorname{argmin}} \|S - uv^T\|_F = \frac{Sv}{\|Sv\|}. \quad (11)$$

Eq. 11 involves multiplication between input matrix S and vector u, followed by setting all elements in the resulting v smaller than its r-th largest value to zero, essentially performing the vector partition operation on v. One rank-1 basis [u, v] can be estimated in each step, afterwards the input matrix S will be deflated to its residual R:

$$R^n = R^{n-1} - uv^T, R^0 = S, 1 < n \leq K, \quad (12)$$

where K is the total number of expected basis (i.e. dictionary atoms) to be discovered from the input data. From a neuroscience perspective, our basic premise is that during both resting-state and task performance, there could be multiple functionally active networks involved in the constitution of the fMRI blood oxygen level dependent (BOLD) signals in the brain. Thus the fMRI signal measured on each voxel is the combination of the co-activations from multiple, yet limited number of functional networks. By assuming such combination is linear, it will then be appropriate and intuitive to subtract the signal on the voxel with the activation patterns from the known networks (e.g. those have been learned in previous steps), with the subtraction weighted by the corresponding contributions of each network, before decomposing new networks.

The total K number of u vectors constitute the learned dictionary matrix U (of dimension $T \times K$). The corresponding sparse vectors v constitute the loading coefficient matrix V (of dimension $K \times P$). The algorithmic pipeline is illustrated in Fig. 18.

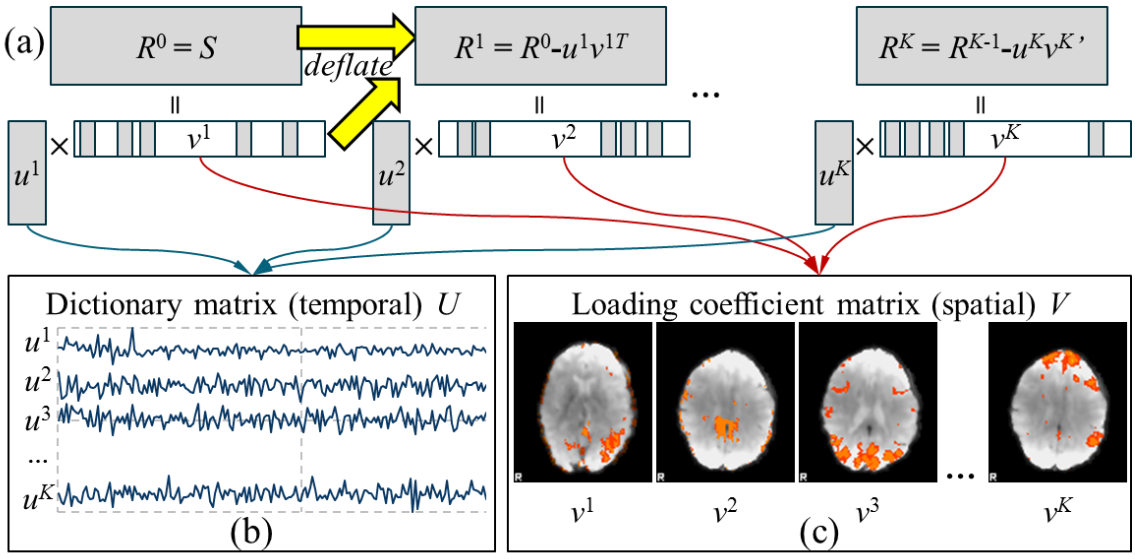

Fig.18. Illustration of the r1DL model applied on fMRI dataset in a running example. (a) Illustration of the iterative deflation procedure during which the input data *S* is decomposed and updated at each step from 1 to *K*. (b) Visualization of the time series in dictionary matrix *U*. (c) Visualization of the corresponding loading coefficient matrix *V*.

As shown in the figure, the input fMRI data can be decomposed into several neuroscientifically meaningful functional networks, with their temporal patterns defined in *U* matrix and spatial patterns defined in *V* matrix. As also observed in previous chapters, *U* matrix has been found to be highly correlated or anti-correlated with stimulus paradigm, while the corresponding *V* matrices were found to be consisted of various task-evoked networks including visual, auditory, motor network and thalamus, anti-task networks including Default Mode Network (DMN), as well as neuroanatomic areas such as the ventricle. To validate the effectiveness of the proposed r1DL model, in this work we applied the model on HCP Q1 release task fMRI (tfMRI) database. The learned results were then compared to the set of functional network atlas identified from the Holistic

Atlases of Functional Networks and Interactions (HAFNI) framework introduced in section 3.2, as well as the 9 resting-state network (RSN) templates reported by applying ICA on rsfMRI data [16]. The comparisons showed that the proposed r1DL model can obtain the similar set of those well-established functional networks using much reduced time with the same parameter settings, as shown in Fig.19. The time cost for decomposing each network is just less than 1 second on a typical machine, thus allows near real-time user feedback. Further, by using smaller dictionary size parameters, r1DL can achieve several-folds speed-up comparing with the previous model.

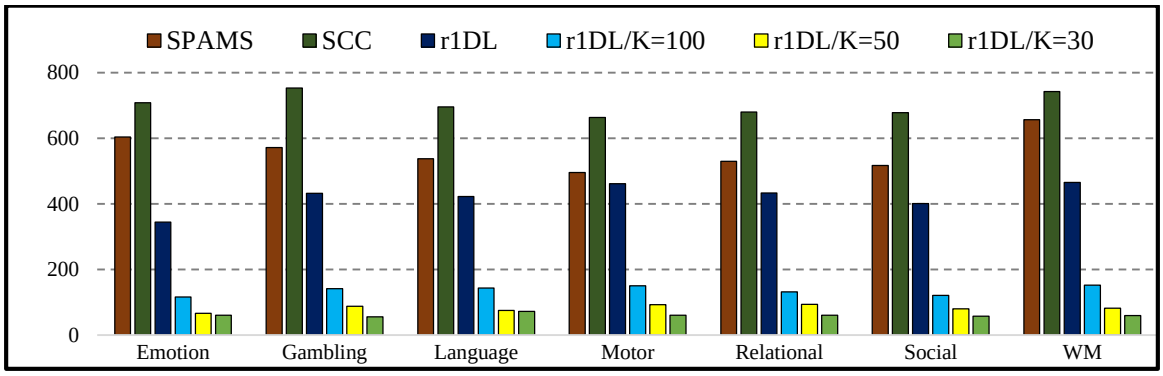

Fig.19. Average time cost (measured in seconds) for decomposing functional networks from the tfMRI data during 7 tasks across 68 subjects using three different dictionary learning methods as well as different dictionary size parameters for r1DLmodel.

In summary, comparing with the whole-matrix gradient-based dictionary learning algorithms, such as the online dictionary learning [75] (based on stochastic gradient descent) and the K-SVD [78] (based on gradient descent), there are several key advantages of the proposed method that could greatly improve the speed and scalability of the network decomposition process. 1) The method does not rely on gradient computation. Thus it does not need to tune the learning rate/step size, and also avoids the slow convergence near the

solution shared by many gradient-based algorithms. 2) Because the method estimates one rank-1 basis vector then saves and discards them at each learning step, r1DL is faster and more scalable especially on large datasets as it avoids maintaining the potentially very large learned results in the memory. 3) From our pilot tests, it was observed that networks learned in the earlier steps were more functionally meaningful, while latter-learned networks tend to be noises. This feature, combining with the fact that in r1DL networks are learned iteratively, helps us to determine the optimal dictionary size, which is a difficult task when applying dictionary learning in practice and a major factor affecting the learning speed. For r1DL method, we can safely set a sufficiently large dictionary size for the iterative learning and truncate the latter noise network components. In addition, the preliminary results have shown that r1DL can obtain similar set of results with smaller dictionary size comparing with the previous methods, thus improve its efficiency for application. 4) The method enables significantly accelerated user feedback and interactive visual analytics, as it can generate the results back to the user based on only the first few functional networks learned, then update the results along with the learning process.

CHAPTER 4

POPULATION-WISE FUNCTIONAL SPACE FOR BIG DATA ANALYTICS ON FUNCTIONAL NEUROIMAGING

4.1 fMRI Big Data Analytics

The models we have developed so far as described through Chapter 2 to Chapter 3 have covered most of the most advanced methods in characterizing functional brain and its dynamics, and have generated important discoveries. However, the models and discoveries we have obtained so far are either subject or group specific, while it has been observed that individual-level or small group-level analysis has the problem of lacking the representability in their results. The problems are generally indicated by the large cross-study variability as well as the lowered reproducibility from their conclusions, possibly due to the fact that the dataset in consideration was too small to sufficiently cover all the brain functional dynamics in the population thus lack the generalizability for the solid and holistic conclusion. For example, the dictionary learning methods are generally applied on individual voxel-wise [9, 10] data, while the connectomics estimation is performed on the regional (DICCCOL) level [3, 8], with the total data size varied from MB (megabyte) to GB (gigabyte) level.

On the other hand, population-level functional network analysis can intrinsically overcome such limitations and provide an important and powerful tool for cognitive and clinical neuroscience studies. The common and consistent functional networks identified across many individuals by the population-wise network analysis can offer a holistic

reference system for studying brain response to various stimulus. In recognition of such importance, studies on population-level network analysis modeling have been getting more focus in the field, including earlier works by [79] who has suggested using a two-stage spatial principle component analysis (PCA) framework to first perform dimension reduction on individual dataset, followed by performing the full component analysis on the aggregated reduced data. Recently, works of [77] used a similar but more robust approach to perform spatial PCA in the population-level analysis of Human Connectome Project (HCP) grayordinates fMRI data from 1200 subjects. It is still a challenging yet increasingly important problem to performing comprehensive network/component analysis on extremely-large and ever-growing dataset with high scalability. Specifically, the desired framework is supposed to include the following features: 1) being scalable to the data size without extra hardware (e.g. memory) requirement; 2) running in parallel/distributed manner, to achieve linear/sub-linear time cost with regarding to data size; 3) obtaining both individual-level and population-level results, the results shall be integrated and comprehensive. In other words, the functional networks obtained from each individual dataset by the framework shall be the same comparing to results obtained by other methods from the same dataset; 4) establishing the correspondence across individual-level results to enable cross-subject, cross-group analysis.

4.2 Distributed Rank-1 Dictionary Learning (D-r1DL) Framework

Following the previous success in using dictionary learning, especially the r1DL model for functional network decomposition and in response to the challenges on fMRI big data analytics listed above, we then devolved the novel distributed rank-1 dictionary learning

(D-r1DL) model, leveraging the power of distributed computing for handling large-scale fMRI big data while at the same time taking advantage of the flexibility of the r1DL model. As the r1DL algorithm is very light-weighted regarding to the operational complexities most of its routines will only take one vector as input and one vector as output. This feature helps the r1DL algorithm to be easily parallelized to its distributed version. We then used the Spark engine to implement the D-r1DL algorithm, which is a high-performance distributed compute engine for large-scale data processing. It is similar to MapReduce, but has several distinct advantages that make it ideal for the deployment of large-scale analytics frameworks. First, its basic abstraction for distributed data, the resilient distributed dataset (RDD) [80], combines robust fault-tolerance with highly efficient data layout strategies. RDDs track their computation lineage as a directed acyclic graph; therefore, if a segment is lost, it can be easily recomputed from the lineage. These lineages can be optimized on-the-fly to minimize the overhead of the prescribed computations. Second, all operations in Spark are performed in-memory, thus significantly improving throughput of data pipelines. This is a departure from Hadoop MapReduce, in which data are serialized to disk in between map and reduce steps. Third, the Spark compute engine is much more generalizable than MapReduce, and can efficiently support highly diverse workloads. While Spark supports the map and reduce primitives from Hadoop MapReduce, it also supports graph processing [81] and streaming [82] APIs on the same compute engine, in addition to numerous functional primitives beyond map and reduce. This structural flexibility is crucial to the efficient implementation of a wide variety of distributed algorithms.

An illustration for the operational and algorithmic pipeline consisting of three layers of model specification is shown in Fig. 20. The first and foremost deliverable of this work is to provide an integrated solution for the large-scale fMRI big data analysis. Therefore, we initially deployed the proposed D-r1DL model on our in-house server (termed "in-house solution") with an integrated neuroinformatics system [83]. The neuroinformatics system provides a web-based user interface for fMRI data uploading, hosting and result post-processing as illustrated in Fig. 20(a). Alternatively, we also tested deploying the D-r1DL model on the cloud computing service provided by Amazon Web Services Elastic Compute Cloud (AWS-EC2), which has been widely applied for biomedical imaging researches due to its resource flexibility and ease of use. For the "AWS-EC2 solution", the data preprocessing was performed before running the D-r1DL model on it. Subroutines of the r1DL algorithm and its logic flow are illustrated in Fig. 20(b). The parallelization subroutines and its relationship with the r1DL algorithm are illustrated in Fig. 20(c).

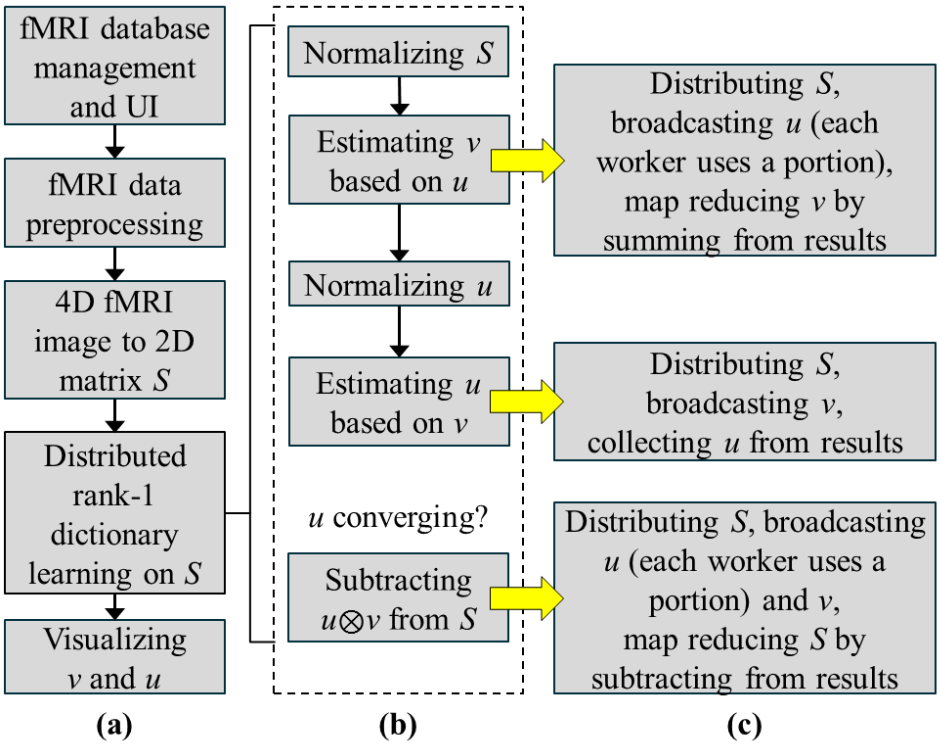

Fig.20. (a): Operations on the neuroinformatics system for preparing the application of D-r1DL and post-analysis. (b): Algorithm pipeline of the rank-1 dictionary learning. (c): Parallelization subroutines of the D-r1DL model derived from the corresponding subroutines of the r1DL using Spark. The distribution of input data S is based on RDDs.

For the parallelization on the Spark engine, we implemented the vector-matrix multiplication and the matrix-vector multiplication steps by their corresponding distributed primitives in Spark. Reading and partitioning the input data *S* is supported by the RDD abstraction; therefore, the distribution of *S* to each node as a series of key-value pairs is inherently straight forward: data formation of the current work is based on row-vectors. In other words, each column in *S* contains the *T* number of observations for one specific feature, to the total of *P* features. While *S* was maintained as an RDD, the vectors *u* and *v* were broadcast to all nodes. Thus during the vector-matrix multiplication, each node will use its portion of the updated *u* vector, and then estimate the *v* vector based on the multiplication of portions of *S* and *u*. The resulting *v* vectors from all the nodes will be then map-reduced by the summation operation. The matrix-vector multiplication is relatively easier, where each node will use all the updated *v* vector then estimate its corresponding portion of the *u* vector. The resulting *u* vector is just the collection of the results from each node. In addition, the matrix deflation operation was also parallelized by broadcasting both the *u* and *v* vectors then estimating the outer produce between portion of *u* vector and the whole *v* vector at each node. The *S* matrix is then subtracted by the results of each node through mapping over each row and deflating it in parallel. Because the distributed primitives added for the parallelization in D-r1DL will potentially cause large extra

computations and/or data transfers across nodes, we analyzed the extra complexity induced by the parallelization of the three subroutines. Assuming that there are *M* number of nodes:

- For the vector-matrix multiplication for estimating v, the total complexity is ($T*M+P*M+T\log(T)+P$): $T*M$ caused by the broadcasting, $P*M+T\log(T)$ caused by the map reduce and network shuffle, and P caused by the updating of *v*.
- For the matrix-vector multiplication, the total complexity is ($P*M+T$): $P*M$ caused by the broadcasting, and T caused by the updating of *u*.
- For the matrix deflation, the total complexity is ($P*M+T*M$): both *u* and *v* will be broadcasted to all M nodes.

The D-r1DL model was deployed on two different sets of server clusters, leading to two solutions for the data analysis. The illustrative diagram showing the organization and execution architecture of the two solutions are shown below.

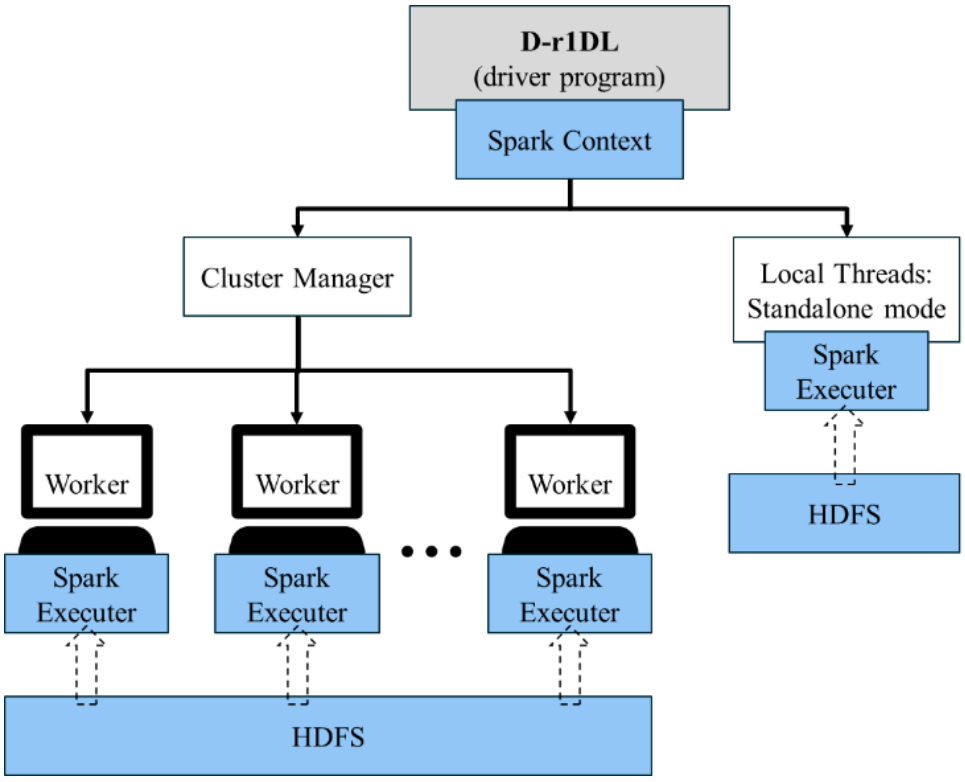

Fig.21. Illustrative diagram showing the organization and execution architectures for the standalone local mode and the multi-worker cluster mode.

Both of the solutions have shown improved performance over the original r1DL model, especially when the dataset is large. In our experiment, the testing input files sizes of the three types of dataset were 300MB, 700MB and 2GB, respectively. Using the in-house solution, we firstly analyzed the performance of the D-r1DL model using different numbers of cores on a single machine. It was found that for all the three datasets, there exists clear logarithmic relationship (R2=0.84, 0.89 and 0.92) between the number of cores recruited and the total time cost for the decomposition. The speed boosts by recruiting more cores for the computation comparing with the baseline (1-core) configuration for the three datasets using the in-house solution are listed in Table 2, showing the ratio between the time cost using 1 core and the time cost using multiple cores. As the configuration for using only one core for D-r1DL is equivalent to the non-parallel algorithm, the performance statists indicate that the parallelization based on Spark could greatly improve the performance of the rank-1 dictionary learning algorithm.

	Emotion	**WM**	**RS**
2 cores	3.1	2.8	1.8
4 cores	6.0	5.1	3.3
8 cores	6.6	7.7	6.3
16 cores	6.8	8.6	6.7

Table.2. Ratios of time cost decreases by recruiting more cores comparing with the single-core configuration.

In addition to the experiments of the single-machine multi-core configurations conducted using the in-house solution, we have also applied the D-r1DL on the same

datasets using the cloud computing service provided by AWS-EC2. We aimed to investigate the performance of D-r1DL when applied over multiple machines through a network interface. Specifically, as the Spark Python architecture and the resilient distributed dataset abstracts have been designed for supporting large-scale, high efficient analytic framework, we are interested to test its capability of utilizing the distributed computational resources provided by AWS-EC2. We tested the performance in terms of time and memory cost of the D-r1DL model using 1, 2, 4, 8 and 16 workers on three datasets, while each worker has two cores for the computation. The D-r1DL would be running in stand-alone mode under single-worker configuration, similar to the configuration used in the in-house solution. As discussed in the complexity analysis aboe, the communications through network interfaces caused by the parallelization of computation (e.g. the broadcasting of u and v) will potentially increase the time cost mainly due to latencies. Thus the single-worker configuration serves as the baseline for testing whether recruiting more workers will be beneficial from the performance perspective. The results shown that the AWS-EC2 solution recorded faster computation speed (10%~80% faster) comparing with the in-house solution, especially on larger dataset, when both of them use two cores. Considering the fact that the hardware configuration of AWS-EC2 features larger memory capacity better optimized for computation purposes, such difference in performance is within our expectation. On the other hand, it is interesting to observe that for AWS-EC2 solution, there exists the break-even point at which the multiple-worker mode outperformed the stand-alone mode, but only for the two larger datasets. For the 700MB WM and the 2GB RS dataset, using 4 or more workers could lead to faster speed comparing with the standalone mode using 1 worker. While for the smaller

300MB Emotion data, the standalone mode is the fastest among all experiments. Thus it can be concluded that the multi-worker configuration will be more suitable for analyzing larger datasets, while standalone mode or the simpler in-house server solution might be preferred for datasets with typically smaller sizes. More importantly, the memory cost on each worker is near constant regardless of the data size, indicating that the multi-worker mode under AWS-EC2 solution scales good with the increasing input file size, as it maintains reasonable small (~100MB) memory cost for all configurations including the single-worker standalone mode. That is the major advantage of using Spark Python model and its resilient distributed dataset for the parallelization: one or multiple workers need not to load the whole dataset at once, but only its corresponding portion of the data according to the data partitioning strategy implemented in the RDDs abstract.

In summary, the better performance gain on larger dataset indicates that the parallelization of the rank-1 dictionary learning could potentially overcome the computational bottleneck for analyzing big neuroimaging data, potentially enabling high-throughput analysis on a locally-deployed high-performance computation cluster in the future. D-r1DL has shown that it is a suitable solution for fMRI big data analytics:

1) **Accuracy**: D-r1DL can discover the same set of results by the General Linear Model (GLM), as well as other functional networks reported in literature such as the well-known resting-state networks (RSNs) also from tfMRI data.

2) **Speed**: D-r1DL distributes computational loads to many nodes, thus achieving greater speed increases with larger clusters. On individual data, the decomposition results are visualized and fed to the user in real-time.

3) **Scalability**: D-r1DL has near-constant memory cost regardless of the input data size as the nodes work on partitions of data rather than the whole dataset. Spark's basic distributed data abstraction, the resilient distributed dataset (RDD), is designed to scale gracefully with the size of the data. In addition, the memory cost of the learning process is minimized to only two vectors.

4) **Deployment**: D-r1DL has been integrated into our in-house neuroinformatics system and is currently running as a publicly-available web service.

Bibliography

[1] Li, X., Lim, C., Li, K., Guo, L., Liu, T., Detecting Brain State Changes via Fiber-Centered Functional Connectivity Analysis, Neuroinformatics, 11(2):193-210, 2013

[2] Zhang, X., Guo, L., Li, X., Zhang, T., Zhu, D., Li, K., Chen, H., Lv, J., Jin, C., Zhao, Q., Li, L., Liu, T., Characterization of task-free and task-performance brain states via functional connectome patterns, Medical Image Analysis, 17(8):1106-1122, 2013

[3] Li, X., Zhu, D., Jiang, X., Jin, C., Zhang, X., Guo, L., Zhang, J., Hu, X., Li, L., Liu, T., Dynamic functional connectomics signatures for characterization and differentiation of PTSD patients, Human Brain Mapping, 35(4):1761-1778, 2014

[4] Ou, J., Lian, Z., Xie, L., Li, X., Wang, P., Hao, Y., Zhu, D., Jiang, R., Wang, Y., Chen, Y., Zhang, J., Liu, T., Atomic dynamic functional interaction patterns for characterization of ADHD, Human Brain Mapping, 35(10):5262-5278, 2014

[5] Ou, J., Xie, L., Jin, C., Li, X., Zhu, D., Jiang, R., Chen, Y., Zhang, J., Li, L., Liu, T., Characterizing and Differentiating Brain State Dynamics via Hidden Markov Models, Brain Topogr, 1-14, 2014

[6] Ou, J., Xie, L., Li, X., Zhu, D., Terry, D., Puente, A.N., Jiang, R., Chen, Y., Wang, L., Shen, D., Zhang, J., Miller, L.S., Liu, T., Atomic connectomics signatures for characterization and differentiation of mild cognitive impairment, Brain Imaging and Behavior, 1-15, 2014

[7] Zhang, X., Li, X., Jin, C., Chen, H., Li, K., Zhu, D., Jiang, X., Zhang, T., Lv, J., Hu, X., Han, J., Zhao, Q., Guo, L., Li, L., Liu, T., Identifying and Characterizing Resting

State Networks in Temporally Dynamic Functional Connectomes, Brain Topogr, 27(6):747-765, 2014

[8] Zhang, J., Li, X., Li, C., Lian, Z., Huang, X., Zhong, G., Zhu, D., Li, K., Jin, C., Hu, X., Han, J., Guo, L., Hu, X., Li, L., Liu, T., Inferring functional interaction and transition patterns via dynamic bayesian variable partition models, Human Brain Mapping, 35(7):3314-3331, 2014

[9] Lv, J., Jiang, X., Li, X., Zhu, D., Chen, H., Zhang, T., Zhang, S., Hu, X., Han, J., Huang, H., Zhang, J., Guo, L., Liu, T., Sparse representation of whole-brain fMRI signals for identification of functional networks, Medical Image Analysis, 20(1):112-134, 2015

[10] Lv, J., Jiang, X., Li, X., Zhu, D., Zhang, S., Zhao, S., Chen, H., Zhang, T., Hu, X., Han, J., Ye, J., Guo, L., Liu, T., Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function, Biomedical Engineering, IEEE Transactions on, 62(4):1120-1131, 2015

[11] Zhang, S., Li, X., Lv, J., Jiang, X., Guo, L., Liu, T., Characterizing and differentiating task-based and resting state fMRI signals via two-stage sparse representations, Brain Imaging and Behavior, 1-12, 2015

[12] Eklund, A., Nichols, T.E., Knutsson, H., Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates, Proceedings of the National Academy of Sciences, 2016

[13] Chang, C., Glover, G.H., Time–frequency dynamics of resting-state brain connectivity measured with fMRI, NeuroImage, 50(1):81-98, 2010

[14] Zhu, D., Li, K., Guo, L., Jiang, X., Zhang, T., Zhang, D., Chen, H., Deng, F., Faraco, C., Jin, C., Wee, C.-Y., Yuan, Y., Lv, P., Yin, Y., Hu, X., Duan, L., Hu, X., Han,

J., Wang, L., Shen, D., Miller, L.S., Li, L., Liu, T., DICCCOL: Dense Individualized and Common Connectivity-Based Cortical Landmarks, Cerebral Cortex, 23(4):786-800, 2013

[15] Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., Correspondence of the brain's functional architecture during activation and rest, Proceedings of the National Academy of Sciences, 106(31):13040-13045, 2009

[16] Smith, S.M., Miller, K.L., Moeller, S., Xu, J., Auerbach, E.J., Woolrich, M.W., Beckmann, C.F., Jenkinson, M., Andersson, J., Glasser, M.F., Van Essen, D.C., Feinberg, D.A., Yacoub, E.S., Ugurbil, K., Temporally-independent functional modes of spontaneous brain activity, Proceedings of the National Academy of Sciences, 109(8):3131-3136, 2012

[17] Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., The WU-Minn Human Connectome Project: An overview, NeuroImage, 80(0):62-79, 2013

[18] Wang, L., Zang, Y., He, Y., Liang, M., Zhang, X., Tian, L., Wu, T., Jiang, T., Li, K., Changes in hippocampal connectivity in the early stages of Alzheimer's disease: Evidence from resting state fMRI, NeuroImage, 31(2):496-504, 2006

[19] Dickerson, B.C., Sperling, R.A., Large-Scale Functional Brain Network Abnormalities in Alzheimer’s Disease: Insights from Functional Neuroimaging, Behavioural Neurology, 21(1-2):2009

[20] Lynall, M.-E., Bassett, D.S., Kerwin, R., McKenna, P.J., Kitzbichler, M., Muller, U., Bullmore, E., Functional Connectivity and Brain Networks in Schizophrenia, The Journal of Neuroscience, 30(28):9477-9487, 2010

[21] Liu, T., A few thoughts on brain ROIs, Brain Imaging and Behavior, 5(3):189-202, 2011

[22] Lindquist, M.A., Waugh, C., Wager, T.D., Modeling state-related fMRI activity using change-point theory, NeuroImage, 35(3):1125-1141, 2007

[23] Robinson, L.F., Wager, T.D., Lindquist, M.A., Change point estimation in multi-subject fMRI studies, NeuroImage, 49(2):1581-1592, 2010

[24] Hu, X., Guo, L., Zhang, D., Li, K., Zhang, T., Lv, J., Han, J., Liu, T.: Assessing the dynamics on functional brain networks using spectral graphy theory. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 2144-2149, 2011

[25] Gilbert, C.D., Sigman, M., Brain States: Top-Down Influences in Sensory Processing, Neuron, 54(5):677-696, 2007

[26] Bassett, D.S., Wymbs, N.F., Porter, M.A., Mucha, P.J., Carlson, J.M., Grafton, S.T., Dynamic reconfiguration of human brain networks during learning, Proceedings of the National Academy of Sciences, 108(18):7641-7646, 2011

[27] Gao, J.-H., Yee, S.-H., Iterative temporal clustering analysis for the detection of multiple response peaks in fMRI, Magnetic Resonance Imaging, 21(1):51-53, 2003

[28] Morgan, V.L., Price, R.R., Arain, A., Modur, P., Abou-Khalil, B., Resting functional MRI with temporal clustering analysis for localization of epileptic activity without EEG, NeuroImage, 21(1):473-481, 2004

[29] Majeed, W., Magnuson, M., Hasenkamp, W., Schwarb, H., Schumacher, E.H., Barsalou, L., Keilholz, S.D., Spatiotemporal dynamics of low frequency BOLD fluctuations in rats and humans, NeuroImage, 54(2):1140-1150, 2011

[30] Sporns, O., Tononi, G., Kötter, R., The Human Connectome: A Structural Description of the Human Brain, PLoS Comput Biol, 1(4):e42, 2005

[31] Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P., Predicting human resting-state functional connectivity from structural connectivity, Proceedings of the National Academy of Sciences, 106(6):2035-2040, 2009

[32] Biswal, B.B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.-M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kötter, R., Li, S.-J., Lin, C.-P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.A.R.B., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.-J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.-C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.-F., Zhang, H.-Y., Castellanos, F.X., Milham, M.P., 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

[33] Hagmann, P., Cammoun, L., Gigandet, X., Gerhard, S., Ellen Grant, P., Wedeen, V., Meuli, R., Thiran, J.-P., Honey, C.J., Sporns, O., MR connectomics: Principles and challenges, Journal of Neuroscience Methods, 194(1):34-45, 2010

[34] Zalesky, A., Fornito, A., Bullmore, E.T., Network-based statistic: Identifying differences in brain networks, NeuroImage, 53(4):1197-1207, 2010

[35] Zhu, D., Li, K., Faraco, C.C., Deng, F., Zhang, D., Guo, L., Miller, L.S., Liu, T., Optimization of functional brain ROIs via maximization of consistency of structural connectivity profiles, NeuroImage, 59(2):1382-1393, 2012

[36] Lv, J., Guo, L., Hu, X., Zhang, T., Li, K., Zhang, D., Yang, J., Liu, T.: Fiber-Centered Analysis of Brain Connectivities Using DTI and Resting State FMRI Data. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part II, pp. 143-150. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)

[37] Faraco, C.C., Unsworth, N., Langley, J., Terry, D., Li, K., Zhang, D., Liu, T., Miller, L.S., Complex span tasks and hippocampal recruitment during working memory, NeuroImage, 55(2):773-787, 2011

[38] Li, K., Guo, L., Li, G., Nie, J., Faraco, C., Zhao, Q., Miller, L.S., Liu, T.: Cortical surface based identification of brain networks using high spatial resolution resting state FMRI data. In: 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 656-659, 2010

[39] Hu, X., Li, K., Han, J., Hua, X., Guo, L., Liu, T., Bridging the Semantic Gap via Functional Brain Imaging, IEEE Transactions on Multimedia, 14(2):314-325, 2012

[40] Passingham, R.E., Stephan, K.E., Kotter, R., The anatomical basis of functional localization in the cortex, Nat Rev Neurosci, 3(8):606-616, 2002

[41] Zhang, T., Guo, L., Li, K., Jing, C., Yin, Y., Zhu, D., Cui, G., Li, L., Liu, T., Predicting Functional Cortical ROIs via DTI-Derived Fiber Shape Models, Cerebral Cortex, 22(4):854-864, 2012

[42] Ramsey, J.D., Hanson, S.J., Glymour, C., Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al. simulation study, NeuroImage, 58(3):838-848, 2011

[43] Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., Network modelling methods for FMRI, NeuroImage, 54(2):875-891, 2011

[44] Sun, J., Hu, X., Huang, X., Liu, Y., Li, K., Li, X., Han, J., Guo, L., Liu, T., Zhang, J., Inferring consistent functional interaction patterns from natural stimulus FMRI data, NeuroImage, 61(4):987-999, 2012

[45] Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., A default mode of brain function, Proceedings of the National Academy of Sciences, 98(2):676-682, 2001

[46] Greicius, M.D., Srivastava, G., Reiss, A.L., Menon, V., Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI, Proceedings of the National Academy of Sciences of the United States of America, 101(13):4637-4642, 2004

[47] Fox, M.D., Raichle, M.E., Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging, Nat Rev Neurosci, 8(9):700-711, 2007

[48] Sabatinelli, D., Lang, P.J., Bradley, M.M., Costa, V.D., Keil, A., The Timing of Emotional Discrimination in Human Amygdala and Ventral Visual Cortex, The Journal of Neuroscience, 29(47):14864-14868, 2009

[49] Van Dijk, K.R.A., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L., Intrinsic Functional Connectivity As a Tool For Human Connectomics: Theory, Properties, and Optimization, Journal of Neurophysiology, 103(1):297-321, 2010

[50] Williams, R., The human connectome: just another 'ome?, The Lancet Neurology, 9(3):238-239,

[51] Kennedy, D.N., Making Connections in the Connectome Era, Neuroinformatics, 8(2):61-62, 2010

[52] Lehmann, D., Michel, C.M., Pal, I., Pascual-marqui, R.D., Event-Related Potential Maps Depend on Prestimulus Brain Electric Microstate Map, International Journal of Neuroscience, 74(1-4):239-248, 1994

[53] Lehmann, D., Strik, W.K., Henggeler, B., Koenig, T., Koukkou, M., Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. Visual imagery and abstract thoughts, International Journal of Psychophysiology, 29(1):1-11, 1998

[54] Pascual-Marqui, R.D., Michel, C.M., Lehmann, D., Segmentation of brain electrical activity into microstates: model estimation and validation, IEEE Transactions on Biomedical Engineering, 42(7):658-665, 1995

[55] Koenig, T., Prichep, L., Lehmann, D., Sosa, P.V., Braeker, E., Kleinlogel, H., Isenhart, R., John, E.R., Millisecond by Millisecond, Year by Year: Normative EEG Microstates and Developmental Stages, NeuroImage, 16(1):41-48, 2002

[56] Yang, M., Zhang, L., Feng, X., Zhang, D.: Fisher Discrimination Dictionary Learning for sparse representation. In: 2011 International Conference on Computer Vision, pp. 543-550, 2011

[57] Olshausen, B.A., Field, D.J., Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature, 381(6583):607-609, 1996

[58] Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y., Robust Face Recognition via Sparse Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2):210-227, 2009

[59] Fedorenko, E., Duncan, J., Kanwisher, N., Broad domain generality in focal regions of frontal and parietal cortex, Proceedings of the National Academy of Sciences, 110(41):16616-16621, 2013

[60] Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., The human brain is intrinsically organized into dynamic, anticorrelated functional networks, Proceedings of the National Academy of Sciences of the United States of America, 102(27):9673-9678, 2005

[61] Dosenbach, N.U.F., Visscher, K.M., Palmer, E.D., Miezin, F.M., Wenger, K.K., Kang, H.C., Burgund, E.D., Grimes, A.L., Schlaggar, B.L., Petersen, S.E., A Core System for the Implementation of Task Sets, Neuron, 50(5):799-812, 2006

[62] Duncan, J., The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour, Trends in Cognitive Sciences, 14(4):172-179, 2010

[63] Pessoa, L., Beyond brain regions: Network perspective of cognition–emotion interactions, Behavioral and Brain Sciences, 35(03):158-159, 2012

[64] Bullmore, E., Sporns, O., Complex brain networks: graph theoretical analysis of structural and functional systems, Nat Rev Neurosci, 10(3):186-198, 2009

[65] Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J., Statistical parametric maps in functional imaging: A general linear approach, Human Brain Mapping, 2(4):189-210, 1994

[66] Logothetis, N.K., What we can do and what we cannot do with fMRI, Nature, 453(7197):869-878, 2008

[67] Krekelberg, B., Boynton, G.M., van Wezel, R.J.A., Adaptation: from single cells to BOLD signals, Trends in Neurosciences, 29(5):250-256, 2006

[68] Beckmann, C.F., DeLuca, M., Devlin, J.T., Smith, S.M., Investigations into resting-state connectivity using independent component analysis, Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1457):1001-1013, 2005

[69] van den Heuvel, M., Mandl, R., Hulshoff Pol, H., Normalized Cut Group Clustering of Resting-State fMRI Data, PLoS ONE, 3(4):e2001, 2008

[70] Chen, H., Li, K., Zhu, D., Jiang, X., Yuan, Y., Lv, P., Zhang, T., Guo, L., Shen, D., Liu, T., Inferring Group-Wise Consistent Multimodal Brain Networks via Multi-View Spectral Clustering, IEEE Transactions on Medical Imaging, 32(9):1576-1586, 2013

[71] Damoiseaux, J.S., Rombouts, S.A.R.B., Barkhof, F., Scheltens, P., Stam, C.J., Smith, S.M., Beckmann, C.F., Consistent resting-state networks across healthy subjects, Proceedings of the National Academy of Sciences of the United States of America, 103(37):13848-13853, 2006

[72] Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., Kelly, M., Laumann, T., Miller, K.L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A.Z., Vu, A.T., Woolrich, M.W., Xu, J., Yacoub, E., Uğurbil, K., Van Essen, D.C., Glasser,

M.F., Resting-state fMRI in the Human Connectome Project, NeuroImage, 80(144-168, 2013

[73] Calhoun, V.D., Pekar, J.J., Pearlson, G.D., Alcohol Intoxication Effects on Simulated Driving: Exploring Alcohol-Dose Effects on Brain Activation Using Functional MRI, Neuropsychopharmacology, 29(11):2097-2107, 2004

[74] Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y., Robust Face Recognition via Sparse Representation, IEEE Trans. Pattern Anal. Mach. Intell., 31(2):210-227, 2009

[75] Mairal, J., Bach, F., Ponce, J., Sapiro, G., Online Learning for Matrix Factorization and Sparse Coding, J. Mach. Learn. Res., 11(19-60, 2010

[76] Lv, J., Lin, B., Zhang, W., Jiang, X., Hu, X., Han, J., Guo, L., Ye, J., Liu, T.: Modeling Task FMRI Data via Supervised Stochastic Coordinate Coding. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015, vol. 9349, pp. 239-246. Springer International Publishing (2015)

[77] Smith, S.M., Hyvärinen, A., Varoquaux, G., Miller, K.L., Beckmann, C.F., Group-PCA for very large fMRI datasets, NeuroImage, 101(0):738-749, 2014

[78] Aharon, M., Elad, M., Bruckstein, A., K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation, Signal Processing, IEEE Transactions on, 54(11):4311-4322, 2006

[79] Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J., A method for making group inferences from functional MRI data using independent component analysis, Human Brain Mapping, 14(3):140-151, 2001

[80] Zaharia, M., Chowdhury, M., Das, T., Dave, A., Ma, J., McCauley, M., Franklin, M.J., Shenker, S., Stoica, I.: Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing. Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, pp. 2-2. USENIX Association, San Jose, CA (2012)

[81] Gonzalez, J.E., Xin, R.S., Dave, A., Crankshaw, D., Franklin, M.J., Stoica, I.: GraphX: graph processing in a distributed dataflow framework. Proceedings of the 11th USENIX conference on Operating Systems Design and Implementation, pp. 599-613. USENIX Association, Broomfield, CO (2014)

[82] Zaharia, M., Das, T., Li, H., Hunter, T., Shenker, S., Stoica, I.: Discretized streams: fault-tolerant streaming computation at scale. Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles, pp. 423-438. ACM, Farminton, Pennsylvania (2013)

[83] Makkie, M., Zhao, S., Jiang, X., Lv, J., Zhao, Y., Ge, B., Li, X., Han, J., Liu, T., HAFNI-enabled largescale platform for neuroimaging informatics (HELPNI), Brain Inf., 2(4):225-238, 2015