NeuroImage: Clinical 12 (2016) 23–33

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Connectome-scale group-wise consistent resting-state network analysis in autism spectrum disorder

Yu Zhao[a,1], Hanbo Chen[a,1], Yujie Li[a], Jinglei Lv[a,b], Xi Jiang[a], Fangfei Ge[a,b], Tuo Zhang[a,b], Shu Zhang[a], Bao Ge[a,c], Cheng Lyu[a,b], Shijie Zhao[a,b], Junwei Han[b], Lei Guo[b], Tianming Liu[a,*]

[a]Cortical Architecture Imaging and Discovery Lab, Department of Computer Science and Bioimaging Research Center, The University of Georgia, Athens, GA 30605, USA
[b]School of Automation, Northwestern Polytechnical University, Xi'an 710072, China
[c]Shaanxi Normal University, Xi'an, Shaanxi, China

ARTICLE INFO

Article history:
Received 4 May 2016
Accepted 6 June 2016
Available online 07 June 2016

Keywords:
Functional brain network
Resting-state network
fMRI
Volume shape descriptor
Autism spectrum disorder
Sparse representation
Connectomics signature

ABSTRACT

Understanding the organizational architecture of human brain function and its alteration patterns in diseased brains such as Autism Spectrum Disorder (ASD) patients are of great interests. In-vivo functional magnetic resonance imaging (fMRI) offers a unique window to investigate the mechanism of brain function and to identify functional network components of the human brain. Previously, we have shown that multiple concurrent functional networks can be derived from fMRI signals using whole-brain sparse representation. Yet it is still an open question to derive group-wise consistent networks featured in ASD patients and controls. Here we proposed an effective volumetric network descriptor, named *connectivity map*, to compactly describe spatial patterns of brain network maps and implemented a fast framework in Apache Spark environment that can effectively identify group-wise consistent networks in big fMRI dataset. Our experiment results identified 144 group-wisely common intrinsic connectivity networks (ICNs) shared between ASD patients and healthy control subjects, where some ICNs are substantially different between the two groups. Moreover, further analysis on the functional connectivity and spatial overlap between these 144 common ICNs reveals connectomics signatures characterizing ASD patients and controls. In particular, the computing time of our Spark-enabled functional connectomics framework is significantly reduced from 240 hours (C++ code, single core) to 20 hours, exhibiting a great potential to handle fMRI big data in the future.

1. Introduction

Seven decades after initially discovered and reported by Kanner (1943) in the United States and Asperger (1944) in Austria, Autism Spectrum Disorder (ASD) has drawn enormous attention from society and the research community due to the large number of affected children and the complicated pathology of the disease. ASD generally develops at an early age. Children diagnosed with ASD usually have difficulty in social interaction and communication, as well as repetitive patterns of behavior, interests, or activities. According to a recent report, ASD affects 1 out of 68 children aged 8 years in the United States (Anon., 2014).

While years of research have shown ASD to be a highly genetic disorder, an understanding of the corresponding phenotype, i.e., the organizational architecture of human brain function and its alteration patterns in ASD patients' brains, are of great significance to yield new insights on the causes *of the disease* and *to* discover potential treatments. In-vivo functional magnetic resonance imaging (fMRI) offers a unique window to investigate the mechanism of brain function and to identify functional network components of *the* human brain (Friston, 2009; Heeger and Ress, 2002; Li et al., 2009; Logothetis, 2008). By analyzing brain activity during different tasks using fMRI, the neurology logics of ASD patients' symptoms can be studied. For instance, atypical activation patterns in Fusiform Face Area during face recognition tasks suggests functional abnormities during face processing; decreased activation in *the* left Inferior Frontal Gyrus and increased activation in *the* Planum and Temporale *regions* during language tasks indicates that instead of integrating the meanings of individual words into a coherent conceptual structure, ASD subjects pay more attention to the meanings of individual words (Stigler et al., 2011). In addition to the abnormal brain activations during tasks, decreased functional connectivity in the default mode network (DMN) were also identified when *the* brain is in *a* resting state (Kennedy and Courchesne, 2008).

Several previous studies involving whole-brain scale ASD related functional connectivities have been reported (Stigler et al., 2011; Kana et al., 2014; Moseley et al., 2015) in the literature, and some studies analyzed distributed networks of correlated activities between localized

* Corresponding author.
[1] Co-first authors.

http://dx.doi.org/10.1016/j.nicl.2016.06.004

brain regions (Fox and Raichle, 2007; Van Dijk et al., 2010) and employed complex network analysis (Bullmore and Sporns, 2009; Rubinov and Sporns, 2010) to characterize connectivity networks. However, there still seems to lack a group-wise consistent study across whole-brain scale regions on large-scale subjects. Most of the previous ASD fMRI studies focused on individual or a relatively small number of brain regions or networks. Mounting evidence has shown that the human brain functions *are realized* via the interactions of multiple concurrent neural networks, each of which is spatially distributed across specific structural substrates of neuroanatomical areas (Dosenbach et al., 2006; Duncan, 2010; Fedorenko et al., 2013; Fox et al., 2005; Pessoa, 2012), The major challenge to elucidating the mechanism of how functions are altered in ASD patients across the entire brain lies in the enormous difficulty in determining the corresponding brain regions of interest (ROIs) in different brains given the remarkable variability in cortical structures and functions across populations (Liu, 2011). The added complications are the unclear cytoarchitectural boundaries between cortical regions and the dramatic changes in functional brain connectivity due to non-linear properties of the brain (Zhu et al., 2013), making the correspondence establishment between different brains even more challenging. The traditional clustering methods and Independent Component Analysis (ICA) (Beckmann and Smith, 2004; McKeown et al., 1998) can well construct anatomically distinct brain networks.

However, *the* human brain is a highly connected system, thus there is no biological reason for different spatial components to hold independent distribution. A brain area could be involved in multiple functional processes and a functional network could recruit various heterogeneous neuroanatomic areas, even in resting states (Lv et al., 2015a,b).

In our previous studies, we have shown that by using sparse representation (SR) method (Mairal et al., 2010) to decode fMRI signals (Lv et al., 2015a,b), the whole brain can be decomposed into multiple concurrent functional network components. Specifically, by using this method to analyze data from 68 HCP (human connectome project) subjects, 23 task related networks and 9 resting state components were consistently identified from 7 different tasks, namely, Holistic Atlases of Functional Networks and Interactions (HAFNI). This finding suggests that there exists common concurrent functional components across individuals, and more importantly, these functional components can potentially solve the difficulty posed by the brain variability and be employed in unveiling the differences in diseased brains (Lv et al., 2015c).

To further derive common networks and patient-specific networks requires a relatively large amount of patients and controls and thereby large amount of fMRI data. Notably, SR method will decompose brain fMRI signals into hundreds of over-complete dictionary components. This big data challenge poses difficulties in making sensible comparisons among healthy controls and diseased brains and then further deriving group-wise consistent networks that are common to patients and controls. In response, we proposed an effective volumetric network descriptor, named *connectivity map*, to compactly and quantitatively describe spatial patterns of brain network maps and resort to the power of distributed system and implemented a fast, novel framework in *the* Apache Spark (http://spark.apache.org/) environment that can identify group-wise consistent networks featured in ASD patients and controls. The applications of our new methods and systems on the Autism Brain Imaging Data Exchange (ABIDE) dataset achieved promising results.

2. Methods

Fig. 1 summarizes the pipeline of our proposed computational framework. First, by using SR method (Lv et al., 2015a), the whole-brain fMRI data of each subject was decomposed into multiple componenets (Section 2.2). Then, the *connectivity map* was obtained from the spatial map of each network component (Section 2.5). To reduce computational complexity, two *similarity* measurements of different levels of accuracy were employed to characterize how similar two components are. The coarser similarity, measured by similarity between *connectivity map*, is calculated as the preliminary similarity measurement and the pair-wise components overlap rate (Section 2.3) is used as the final similarity measurement. A sparse similarity matrix between all SR components obtained from all subjects was generated using high-speed computing framework implemented on Apache Spark (Section 2.4). Then spectral clustering was performed on the similarity matrix to identify group-wise consistent brain resting-state networks (ICNs) across individuals. For those component clusters that can be reproducibly identified in individuals, *an* ICN template was generated for each of them. Then the interactions between those ICNs were further analyzed in individual brain space. Specifically, a two-sample *t*-test was performed to identify the inter-network interactions that are significantly different between ASD patients and healthy controls as connectomics signatures. In addition, a support vector machine (SVM) classifier was trained to differentiate ASD patients from controls based on the connectomics signatures. To better understand the details of the networks that show differences between controls and paitents, another round of spectral clustering was carried out on the interactions of ICNs.

2.1. Experimental data and preprocessing

Our experimental data was downloaded from the publicly available Autism Brain Imaging Data Exchange (ABIDE) which provides previously collected resting state functional magnetic resonance imaging (rsfMRI) datasets from individuals with ASD and healthy controls (http://fcon_1000.projects.nitrc.org/indi/abide/). In this study, the rsfMRI data obtained from 77 ASD patient individuals and 101 healthy controls from NYU Langone Medical Center were used to develop and test our computational framework. The acquisition parameters were as follows: 240 mm FOV, 33 slices, TR = 2 s, TE = 15 ms, flip angle = 90°, scan time = 6 min, voxel size = 3 × 3 × 4 mm.

Data preprocessing based on FSL (Jenkinson et al., 2012) was similar to that in (Lv et al., 2015a), which includes *skull removal*, motion correction (MCFLIRT command in FSL tools was adopted, where 4 mm (voxel level) motion parameter search were conducted to remove micro head-motions (Jenkinson et al., 2002). Also note that dissociation results between DMN subnetworks were demonstrated very similar with and without motion scrubbing (Starck et al., 2013).), spatial smoothing, temporal pre-whitening, slice time correction, global drift removal, and linear registration to the MNI space. All of these steps are implemented by FSL FEAT and FLIRT. Then the fMRI data was decomposed into different functional networks using sparse representation (SR) (Lv et al., 2015a) introduced later.

2.2. Sparse representation of whole-brain fMRI data

Sparse representation is a useful machine learning method that can faithfully reconstruct the signal and achieve a compact representation of signals. Based on sparse representation, the whole-brain fMRI signals of each individual could be decomposed into multiple network components as proposed in (Lv et al., 2015a). The whole process is illustrated in Fig. 2. First, the BOLD signal in each voxel of fMRI data *was* normalized. Then the normalized signals in the whole brain were extracted from fMRI data to form a matrix $X \in \Re^{t \times n}$ with n columns containing normalized BOLD signals from n foreground voxels. By applying the online dictionary learning and sparse coding method (Mairal et al., 2010), each column in X was modeled as a linear combination of atoms of a learned basis dictionary D such that $X = D \times \alpha + \varepsilon$, where $D \in \Re^{t \times m}$ is a dictionary matrix, $\alpha \in \Re^{m \times n}$ is a sparse coefficient matrix, and ε is the error term. Finally, each row in the α matrix was mapped back to the brain volume as a network component for future analysis. In (Lv et al., 2013), the authors have shown that the meaningful networks

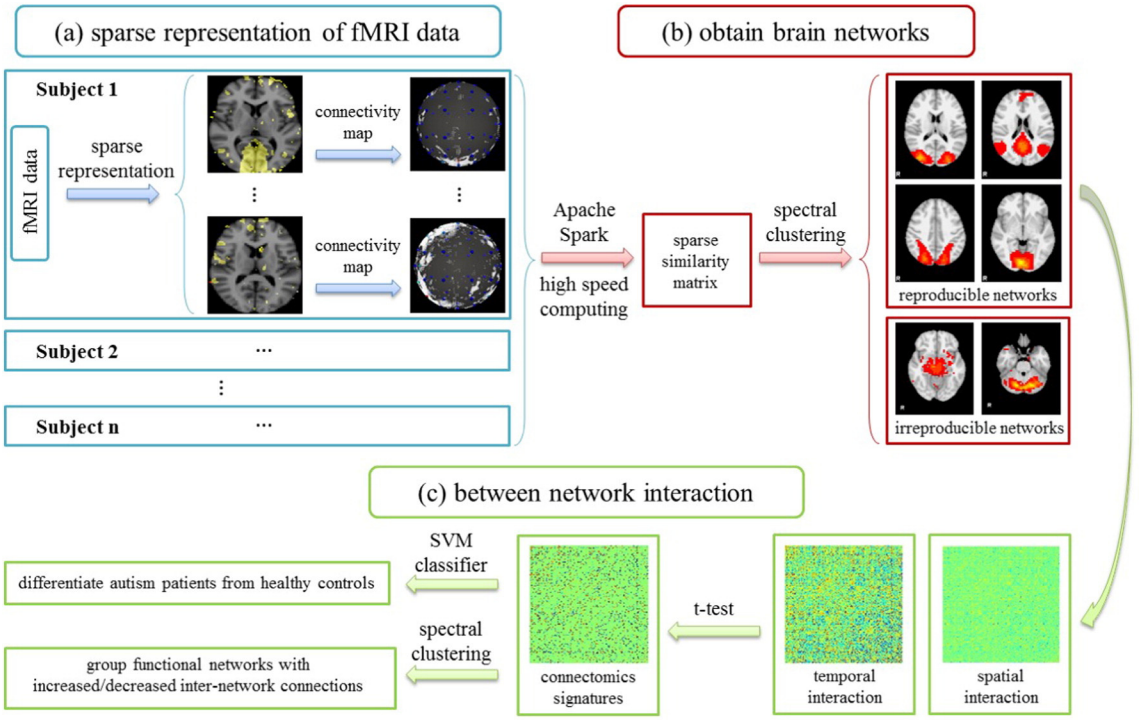

Fig. 1. Flow chart of the proposed computational pipeline.

decomposed do not change significantly with the alteration of dictionary size. Considering the size of fMRI data matrix and the number of signals, *dictionary* size was empirically set to 200 and *sparsity constraint lambda was set to 0.15* for *this study*.

2.3. Brain network similarity based on overlap rate

To identify group-wise common brain networks, it is essential to first define a quantitative measurement of the similarity between brain networks. Since all the subjects have been pre-registered to the MNI space, it is intuitive to use the overlap rate between spatial maps of brain networks as a similarity metric. Specifically, we performed voxel-wise comparisons and define overlap rate similarity (ORS) as the summation of the minimum value in each voxel between two components over the summation of the averaged value in each voxel as the following:

$$S(v_i, v_j) = \frac{\sum_{k=1}^{|V|} \min(v_{i,k}, v_{j,k})}{\sum_{k=1}^{|V|} (v_{i,k} + v_{j,k})/2} \quad (1)$$

where $v_{i,k}$ is the value in the k^{th} voxel of a brain network component V_i. The larger the overlap rate is, the more similar two components are (Fig. 4). The major advantage of this measurement is that it takes into consideration each of the foreground voxels of both components and thereby offers an accurate estimation of the similarity. The drawback,

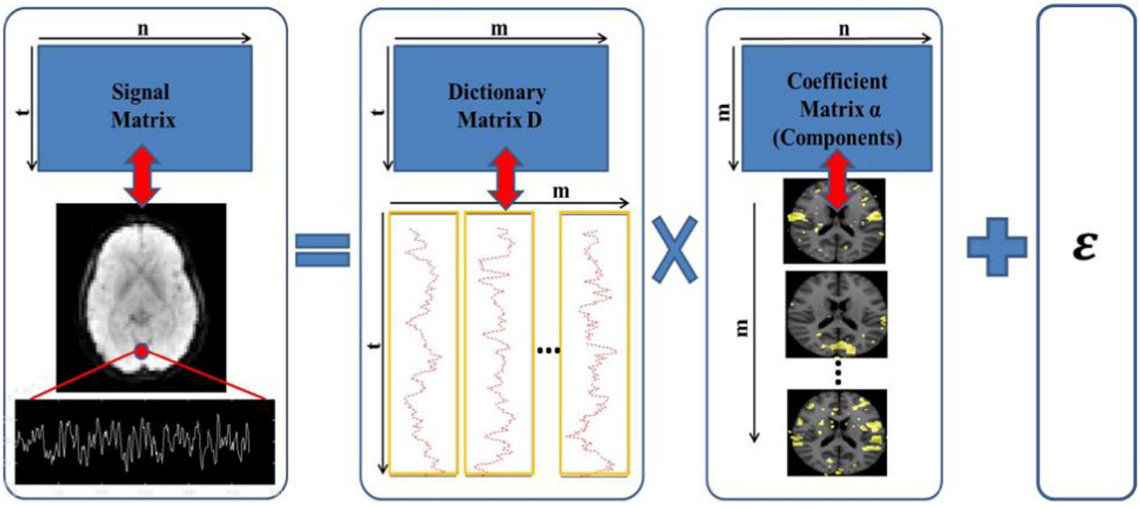

Fig. 2. Illustration of network components generated by sparse representation.

traded with the high accuracy, is the computational complexity. Since there are a total of 35,600 components from 178 subjects and each component has a feature length of over 100,000 (i.e., the number of foreground voxels in a brain image), pairwise similarity is both very time-consuming and memory-expensive *to* compute. Two solutions were proposed and explored here. One is to perform feature dimension reduction without compromising too much on accuracy. The other approach is to employ the power of data-intensive computing platform of Apache Spark. Both strategies will be detailed in the following sections.

2.4. Apache spark implementation for speed-up calculation

Building pair-wise correlation *matrices* requires a polynomial time algorithm with a time complexity of $O(n^2)$, where n is the total number of components (35,600 in this study). Even if there is enough memory to load the volumes of all the components, it will take approximately 1280 h by using single core to compute pair-wise ORS (C++ implementation).

To efficiently manage memory and to employ the power of parallel/distributed computing, the whole computation was implemented on Apache Spark (https://spark.apache.org). Apache Spark is an open-source cluster computing framework originally developed in the AMPLab at *the* University of California Berkeley, which provides a fast and general engine for big data processing. The key steps to deploy Spark in this task include loading each volumetric component as a Resilient Distributed Dataset (RDD), which is the basic distributed memory abstraction in Spark and re-organizing the data to perform pairwise comparison using Cartesian products. The output of the processing is a sparse similarity matrix.

Notably, Apache Spark takes care of job scheduling and resource allocation among multiple clusters and cores. It is a very powerful tool for complex, multi-step data pipelines development. Another useful feature is that Spark not only performs in-memory computing, but also allows data exchange between memory chips and hard drives with optimized speed when the memory is not sufficient to store the big data. Due to these properties, Spark is ideal in solving our big data problem of calculating similarity *matrices* among all of the SR components. With Spark, the original 1280 h computation on one core can be significantly shortened to roughly 100 h running on our 16-core server.

2.5. Brain network similarity based on connectivity map

To further reduce the computational complexity, dimensionality reduction techniques were exploited. Here we proposed a compact volume shape descriptor named *connectivity map*. The basic idea of *the connectivity map* is to unfold the spatial pattern of volumetric voxels by projecting them to points on a unit sphere. Then by sampling the distribution of points on the sphere, a 1-dimensional numerical vector can be obtained to describe the distribution pattern of *the* spatial map. The idea was *inspired by* the shape descriptor of streamline bundles (Chen et al., 2013b; Zhu et al., 2012). But the novelty here is that it is customized for brain network description in the 3D volumetric space. The procedure of projection is illustrated in 2D space as shown in Fig. 3(a)–(b). An example of a 3D volume and obtained *connectivity map* is shown in Fig. 3(d)–(f). The mapping procedure from a network map to a *connectivity map* is listed as follows.

1. Select a projection center v_0 in the 3D space (red point in Fig. 3(a) and (d));
2. Calculate the vectors from the projection center to each foreground voxel in the network map (red arrows in Fig. 3(a)–(b)) and normalize them such that each vector can be represented by a point on unit

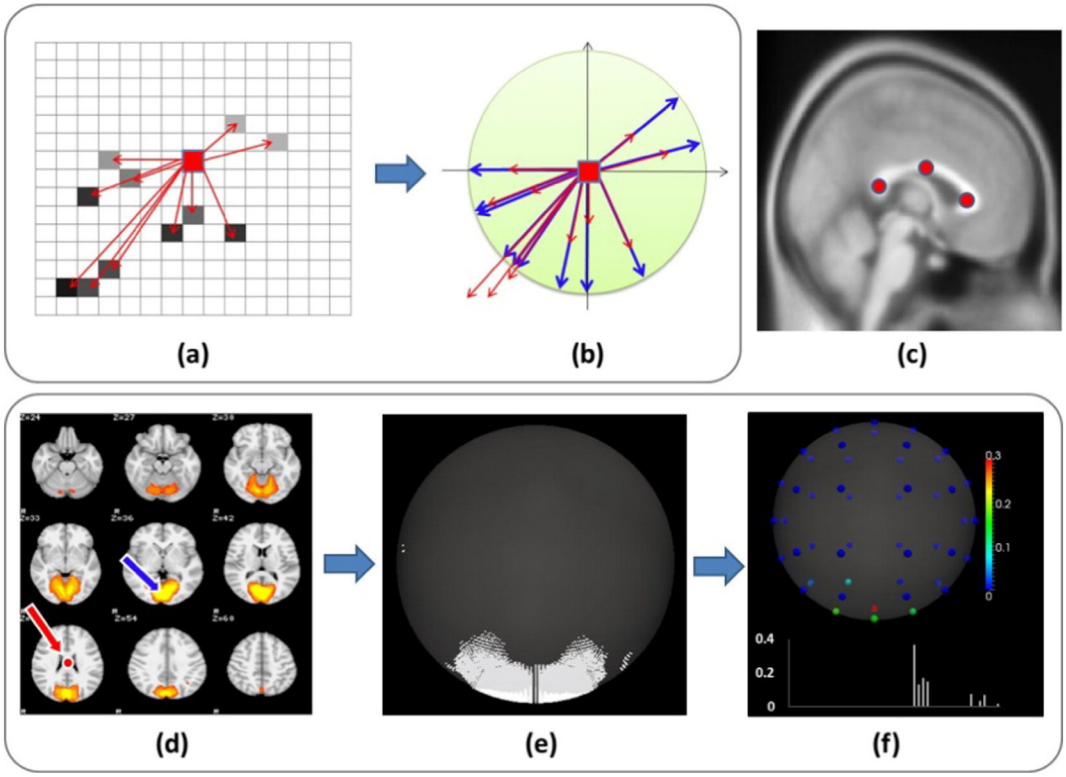

Fig. 3. Illustration of *connectivity map* for volume image. (a) 2-D spatial map; (b) projection of spatial map in (a) to unit circle; (c) 3 selected projection centers represented by red dots; (d) 3-D spatial map of a functional network; (e) projection of spatial map in (c) to unit sphere; (f) *connectivity map* for (c). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sphere (blue arrows in Fig. 3(b)) and its intensity recorded accordingly:

$$W = \left\{ \left(i_k, \overrightarrow{u_k} \right) | \overrightarrow{u_k} = \left(\overrightarrow{v_k} - \overrightarrow{v_0} \right) / |\overrightarrow{v_k} - \overrightarrow{v_0}|, k \in V \right\} \quad (2)$$

where V is the set of a brain network's foreground (non-zero value) voxels; W is the set of projected points; v_k and v_0 are the coordinates of voxels and projection centers; i_k is the voxel intensity (the weighted coefficient in the sparse representation).

3. The probabilistic distribution density of P on the unit sphere is applied to describe the connection map of the voxel data. Specifically, the sphere is segmented into 48 equally sized regions (Chen et al., 2013b). Then, the intensities of projected points in each region are accumulated and normalized:

$$P_j = \sum_{k \in W \cap R_j} i_k / \sum_{k \in W} i_k (j = 1 \cdots 48) \quad (2)$$

where R_j is the area covered by region j, P is the *connectivity map* of a volumetric image.

Considering the property of *the connectivity map* as a probability density vector, the intersection between *connectivity maps* of two networks V_i and V_j can be applied as a similarity measurement:

$$S(P(V_i), P(V_j)) = \sum_{k=1}^{48} \min(P_k(V_i), P_k(V_j)) \quad (3)$$

By this definition, the similarity value will be between 0 and 1, and a higher similarity value indicates networks of higher similarity (Fig. 4).

The description power and the accuracy of *connectivity maps* largely depend on brain orientation and the selection of projection centers. Intuitively, once the brain orientation or the location of projection center changes, the *connectivity map* for the same spatial map may also vary. Thus, to make *connectivity maps* of network spatial maps comparable among individuals, we need to 1) align the brains in the same orientation *and* 2) select projection centers at the same location among individuals' brains. Meanwhile, *connectivity map* may lose description power if the projection center locates inside the network cluster of *a* brain network. For instance, if we select the activation center highlighted by the blue arrow in Fig. 3(d) as the projection center, the projected points will be evenly distributed on the unit sphere and thus it will be hard to tell the pattern of network spatial map with a *connectivity map*. Another limitation of *the connectivity map* is that it cannot identify two different networks when they locate in the same orientation from the perspective of projection center. To solve these issues, as shown by red dots in Fig. 3(c), we selected 3 projection centers along the corpus callosum and generated 3 *connectivity maps* for each network map analyzed. The final *connectivity map* similarity (CMS) is defined *as* the average similarity between 3 pairs of *connectivity map*. Our rational is that: 1) by selecting projection *centers* in white matter regions, the situation that it locates inside network clusters could be eliminated; 2) by applying multiple projection centers to observe each network spatial map from different views, it could be ensured that different network maps will result in different *connectivity maps*; 3) by selecting projection *centers* from consistent and identical anatomical structures in the brain, the errors caused by individual variability can be reduced and *the connectivity map*'s accuracy in comparing brain networks across individuals can be *improved*.

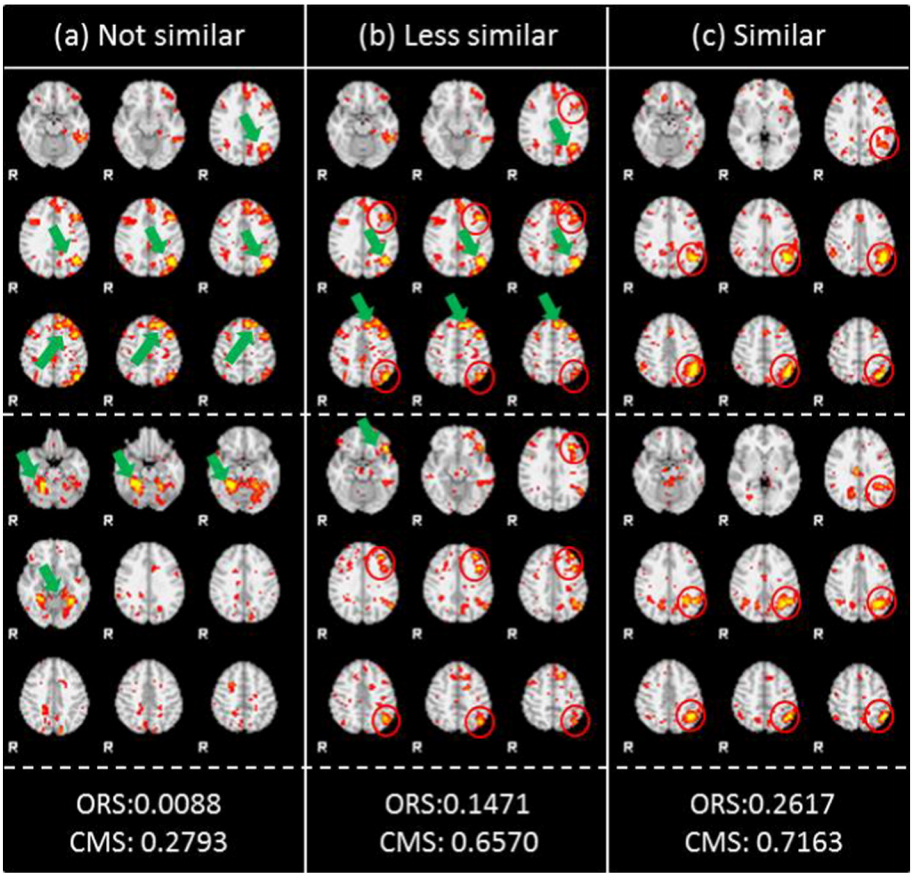

Fig. 4. Examples of similarity measurement. *Connectivity map* similarity (CMS) and overlap rate similarity (ORS) between each pair of selected SR components were listed below each subfigure. The similar parts were highlighted by red circles and dissimilar parts were highlighted by green arrows.

Our testing results showed that CMS performs quite well in identifying dissimilar SR components (Fig. 4(a)). Yet its performance in discerning similar SR components is less robust. In several cases, dissimilar components may also receive a high CMS (Fig. 4(b)–(c)). Nevertheless, given its precise and fast ability in identifying dissimilar SR components, CMS can be employed initially to quickly filter out evidently dissimilar SR components with high confidence. Later, similarities between the remaining components are measured by the more accurate ORS. Intuitively, by increasing CMS threshold, the amount of similarities to be calculated using ORS will be significantly reduced, while the chance of ignoring similar components will be increased. Thus in our computation, a relatively conservative CMS threshold (0.5) was chosen to minimize accuracy trade off during computational speedup. And only when ORS is larger than 0.2, the two components will be considered as similar and the ORS between them will be recorded. After integrating CMS measurement into our computational framework, further speedup was achieved and the whole computation only took 64 h to finish by using 16 cores on a single server.

3. Result

3.1. Resting-state networks in autism and healthy control

By using the method proposed, 35,600 SR components were obtained from the rsfMRI data of 77 ASD patients and 101 healthy controls. The inter components similarity matrix was then calculated. We performed clustering on the obtained similarity matrix to cluster similar SR components. First, before clustering, the components with weak connections (low similarity to other components) were eliminated – if the degree of a component is less than 5 after binarizing the connectivity with 0.25 threshold, the component will be eliminated before clustering. Then, spectral clustering (Chen et al., 2013a; Luxburg, 2007) was performed to iteratively bi-partition the component clusters. Specifically, normalized cut (Chen et al., 2013a; Luxburg, 2007) was applied as the stopping criteria of bi-partition – bi-partition will stop if the corresponding normalized cut value is larger than 0.7. 199 initial clusters (http://hafni.cs.uga.edu/autism/init) were obtained after clustering. Based on our visual inspection, all the clusters obtained were meaningful – similar components were clustered together and clusters were discriminative to each other. Intriguingly, all initial clusters obtained contain SR components from both ASD patients and healthy controls. No clusters are ASD specific or healthy specific.

For each initial cluster, a network template was derived by averaging all the SR components within this cluster. Then we searched the correspondence of each network in the SR components of each individual's brain – the component with the maximum ORS to the templates was taken as its correspondence. Notably, if the ORS between a template and its correspondence is smaller than 0.2 in more than 50% ASD subjects and 50% healthy controls, the network was taken as irreproducible and was eliminated for further analysis. Besides, the motion-induced artifact ICN templates are further removed from analysis. In total, 144 networks remain and form the group-wise common ICNs in our used dataset. For the convenience of further discussion and easy visualization, all of these ICNs were sorted by their reproducibility and each ICN is given an index (http://hafni.cs.uga.edu/autism/templates/all.html).

For each common ICN, we then performed one-tailed *t*-test to compare the ORSs between ASD patients and healthy controls. 4 ICNs with significantly higher (p-value: 0.025) ORS in ASD patients were identified (Fig. 6). 12 ICNs with significantly higher (p-value: 0.025) ORS in healthy controls were also identified (Fig. 7). Multiple comparison test corrections by applying False Discovery Rate (FDR) Control for p-value are performed later to obtain FDR corrected p-value (pFDR-value). And significant results that still remained are shown inside the red boxes in Figs. 6 and 7. Furthermore, for the template 12 with the *t*-test p-value indicates significant difference among two populations, and it appears to be a white matter component or artifact component without physiological meaning. Since the functional disorders in autism patients are not fully understood, this kind of finding can still be addressed for further discussion.

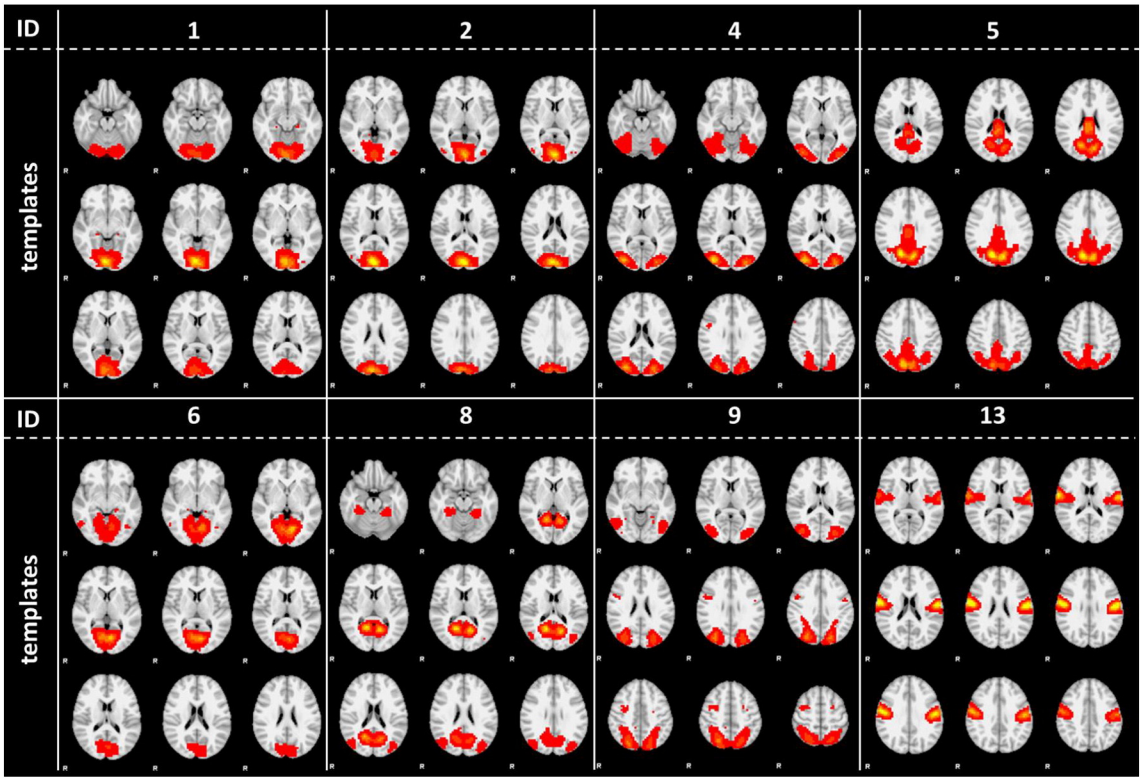

Fig. 5. 8 of 128 common ICNs consistently identified in both autism patients and healthy controls.

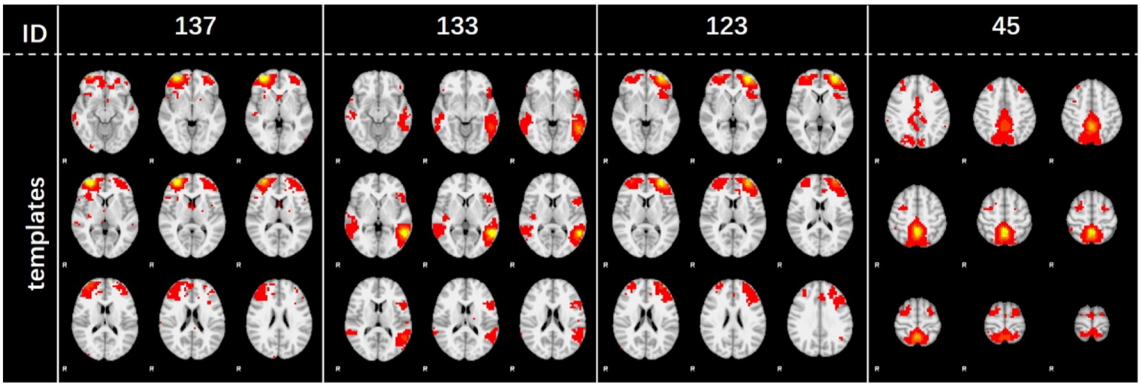

Fig. 6. Templates of ICNs with significantly higher (p-value: 0.025) ORS in ASD patients. All the networks were sorted by ascending p-values. No template of ICNs with significantly higher (pFDR-value) ORS in controls has left after multiple comparison correction.

For the rest of 128 ICNs, they were commonly distributed in both populations (8 of which are shown in Fig. 5). An increased/decreased ORS reflects increased/decreased connections within each ICN. A decreased ORS in DMN (network 7, Fig. 8(a)–(b)) agrees with the previous findings that the connection within DMN decreases in ASD patients in comparison with the healthy controls during resting state

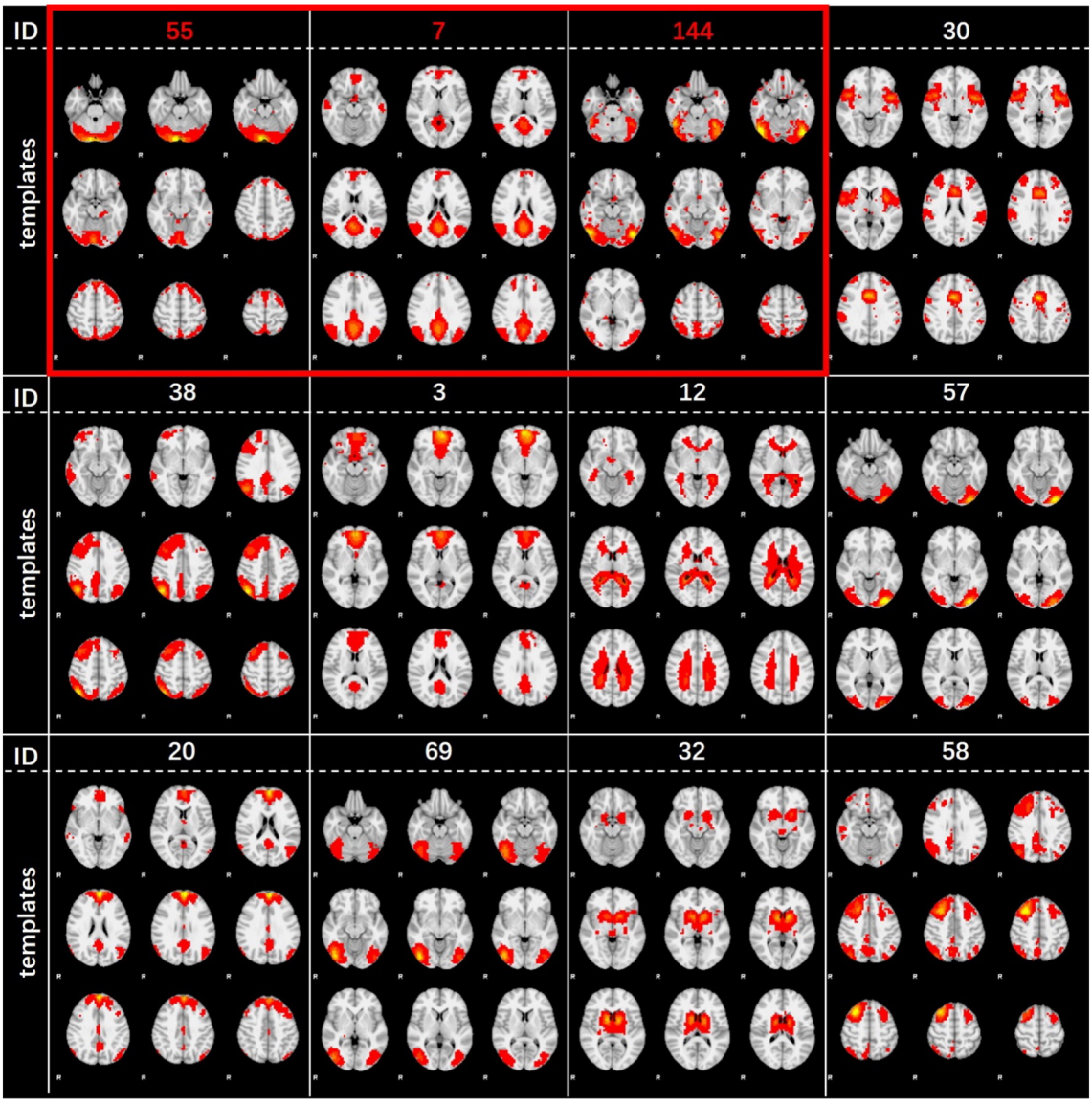

Fig. 7. Templates of ICNs with significantly higher (p-value: 0.025) ORS in healthy controls. All the networks were sorted by p-values. Templates of ICNs with significantly higher (pFDR-value) ORS in patients after multiple comparison correction are listed inside the red box. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Assaf et al., 2010; Kennedy and Courchesne, 2008). In addition, we identified increased connections within Right Frontal Pole (FP) (network 137); Left FP (network 123); Middle Temporal Gyrus (MTG) (network 133); and Precuneous Cortex (PC) (network 45). The functional networks with decreased connections (decreased ORS) we identified includes Occipital Fusiform Gyrus (OFG) / Inferior Lateral Occipital Cortex (ILOC) (network 144, 57, 69); Anterior Cingulate Gyrus (ACG) / Insular (network 30); Right Superior Lateral Occipital Cortex (SLOC) / Right Middle Frontal Gyrus (MFG) (network 38); PC / FP / Frontal Medial Cortex (FMC) / ACG / Posterior Cingulate Gyrus (PCG) (network 3); FP / Superior Frontal Gyrus (SFG) / PC (network 20); Caudate / Putamen / Thalamus (network 32); Right MFG / Right SFG / Right SLOC (network 58). Intriguingly, we also identified decreased functional connections within Cerebral Crus (network 55) and Lateral Ventricle (network 12). We will discuss these findings later.

3.2. Between network interactions

In order to further investigate the differences between autism patients and healthy controls, we computed the interactions between all 144 ICNs in each individual's brain. First, time series were extracted from the dictionary atoms as introduced in Section 2.2. The temporal interactions between each pair of networks are measured using the Pearson correlations of the extracted time series. And the spatial interaction is measured by the ORS between corresponding components. Then based on two-sample *t*-test with 1000 permutations, the interactions that are significantly different (*p*-value: 0.05) between ASD patients and healthy controls were selected as connectomics signatures. Based on these connectomics signatures, SVM classification was performed and tested in 10 folds cross validation manner. On average, we achieved 93.8% accuracy, 91.1% sensitivity, and 95.3% precision in differentiating ASD subjects from the healthy controls, which is quite promising and encouraging.

One-tailed *t*-test was performed to identify the interactions that are significantly increased or decreased in ASD patients. The findings in both types of interactions were largely consistent with each other (Fig. 9). There is no conflict finding between temporal interactions and spatial interactions – if one interaction significantly increased, the other either significantly increased or had no significant differences between ASD patients and healthy controls. All the interactions significantly changed were listed in Fig. 9(c) together with the differences in average temporal interactions (Fig. 9(a)) and average spatial interactions (Fig. 9(b)). Though all of the 144 functional networks can be identified in both populations and most of them have common distributions across subjects, significant alternations in the inter-network interactions were identified among most of the ICNs (Fig. 9). For instance, the temporal interactions of DMN (network 7) (Fig. 8(a)) significantly decreased (p-value: 0.025) with network 3 (PC / FP / FMC / ACG / PCG), 15 (Lingual Gyrus (LG) / PC / Intracalcarine Cortex (IC)), and 50 (FMC / Right Frontal Orbital Cortex (FOC)) (Fig. 8(c)). While its interaction significantly increased (p-value: 0.025) with network 39 (LG / OFG), 76 (Left Angular Gyrus (AG) / Left SLOC / Left FP / Left MFG / Left SFG), 79 (Cerebral Crus / OFG), 84 (Right Posterior Supramarginal Gyrus (PSG) / Right AG / Right FP), 111 (Right MTG / Right AG / Right Supramarginal Gyrus (SG)), 129 (Right Inferior Frontal Gyrus (IFG) / Right FP / FOC) in ASD subjects (Fig. 8(d)). Intriguingly, increasing interactions between DMN and multiple language-related ICNs were identified. This may relate to the impairment in social interactions and communication of ASD subjects and will be further discussed.

Based on the results of one-tailed *t*-test, we further performed clustering to cluster brain ICNs with strong interaction increase or decrease. Specifically, two affinity matrices (increased/decreased) were

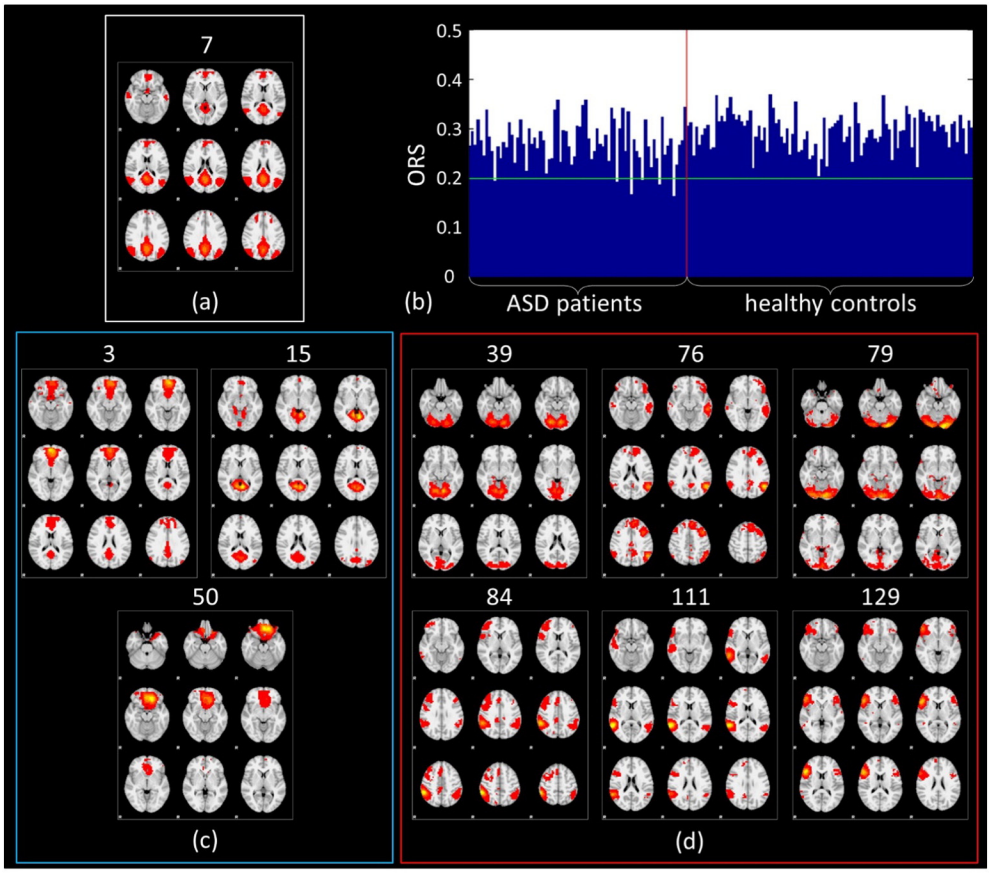

Fig. 8. (a) Spatial map of network 7. (b) ORS between the template of network 7 and the corresponding SR components in each individual's brain. (c) ICNs with decreased temporal interactions to network 7. (d) ICNs with increased temporal interactions to network 7.

generated such that if both temporal and spatial interactions are significantly increased or decreased, the corresponding connectivity is defined as 1; and if only one type of interaction significantly increased or decreased, the connectivity was defined as 0.5, otherwise the connectivity was 0. For each affinity matrix, we first eliminated the ICNs with weak interactions, and then performed spectral clustering to cluster ICNs with strong interactions. 14 network clusters were finally obtained for both increased interactions and decreased interactions, and for convenience, they are referred as increased/decreased network interaction clusters (INICs/DNICs) (Fig. 9). Intriguingly, we also found a considerable amount of decreased interactions within INICs as well as increased interactions within DNICs (highlighted by black arrows in Fig. 9). For interactions within INIC, 27.9% of within cluster network interactions increased while 3.9% decreased. And for interactions within DNIC, 28.4% of within cluster network interactions decreased while 1.7% increased. The complicated changes in functional network interactions might be related to the network alterations and functional compensation.

3.3. Method reproducibility

In order to demonstrate the reproducibility of our method, 152 group-wisely consistent ICNs were obtained by applying the proposed method with the same parameters to rsfMRI datasets from 9 different sites (KKI, NYU, Olin, SBL, Stanford, UCLA, UM, USM, and Yale) on the ABIDE websites. All the 144 ICNs obtained from the NYU site are reproducible in the newly-obtained 152 ICNs from 9 sites, which means that each ICN out of 144 ICNs has a correspondent ICN from the 152 ICNs with a ORS larger than 0.2. Fig. 10 shows the ORS values between the 144 templates and the corresponding templates in the newly-obtained 152 templates, which demonstrated the correspondence between the 144 templates and the newly obtained 152 templates and thus further validated the reasonable reproducibility of our proposed methods. The method reproducibility webpage in our data portal (http://hafni.cs.uga.edu/autism/templates/method_reproducibility.html) showed the

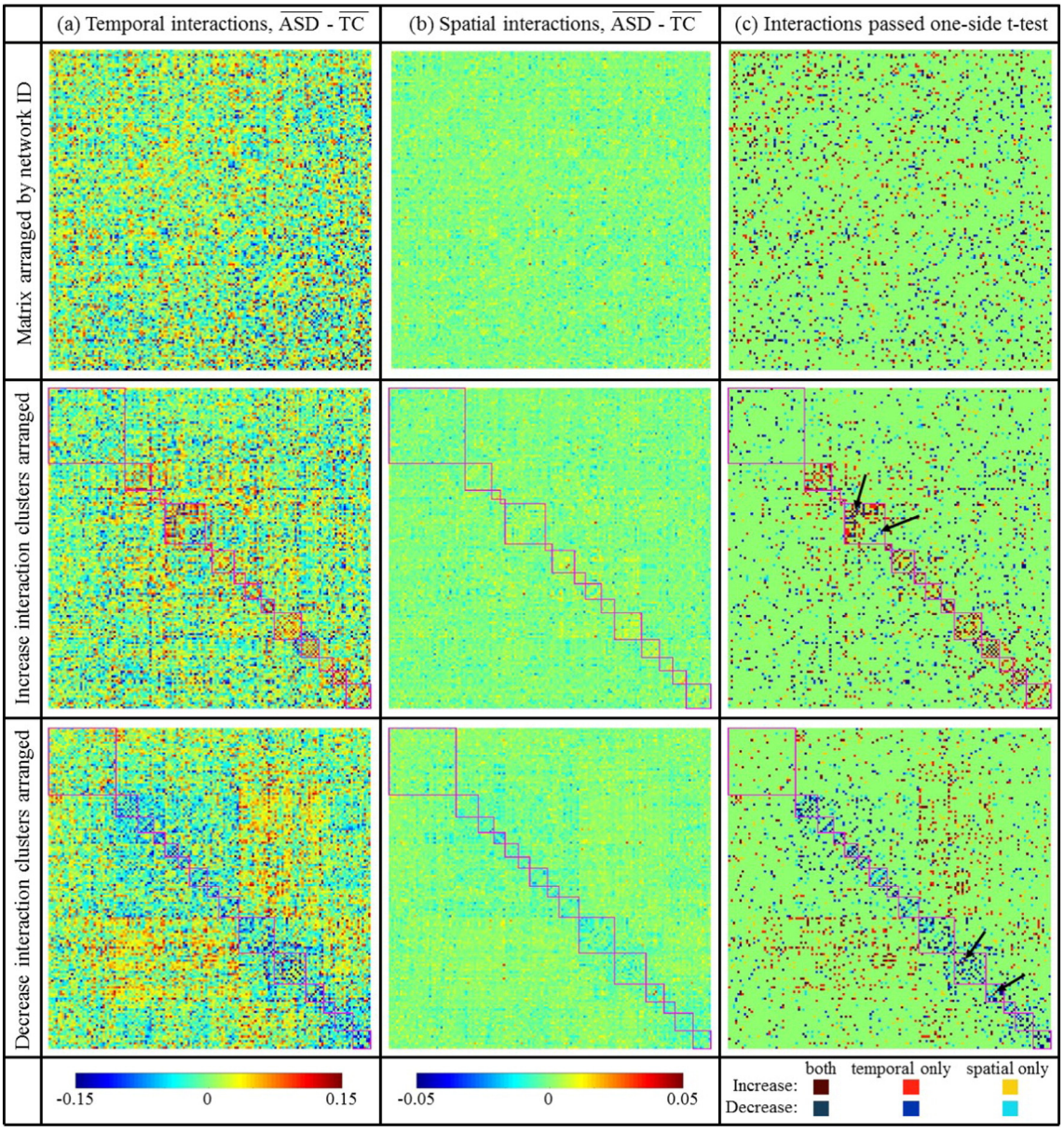

Fig. 9. Group-wise comparisons of inter-network interactions. (a) Differences in average temporal interactions between ASD patients and healthy controls (TCs). (b) Differences in average spatial interactions between ASD patients and TCs. (c) Significantly increased or decreased interactions identified by one-tailed *t*-test (p-value: 0.025). Color bars were listed at the bottom of subfigures. The matrices in the second and third rows were arranged by clusters. In each subfigure, starting from top left, the 1st magenta box is the ICNs with weak connections. Following it, the rest of the boxes are clusters from #1 to #14 respectively.

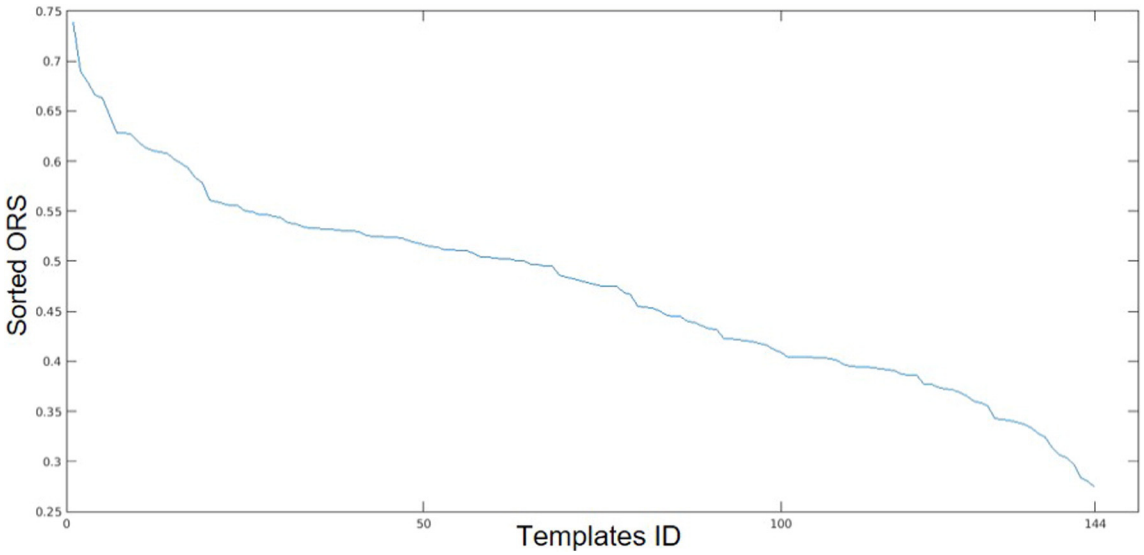

Fig. 10. All ORS values of the corresponding templates from the reproduced 152 templates are greater than 0.2, demonstrating the proposed method is reasonably reproducible on other sites' data.

reasonably good correspondence between the previous generated 144 templates and the independently reproduced 152 templates.

In addition, the connectomics signatures generated from two-sample *t*-tests with 1000 permutations on NYU training dataset were calculated as features from 9 sites' datasets for SVM classifier testing for the feature reproducibility validation. It turned out that a 70.48% 10-fold cross validation accuracy was achieved on datasets from 9 ABIDE sites. This reasonable accuracy suggests that our connectomics signatures are predictive on separate datasets, though the accuracy can be further substantially improved in the future. Meanwhile, except for NYU training dataset, 9 other sites' datasets were used to generate the connectomics signatures using two-sampled t-tests with 1000 permutations and are then calculated for SVM model training, with a 76.46% 10-fold cross validation accuracy was achieved, which demonstrated that the features selected using our proposed method can generalize to other datasets other than only the NYU dataset.

4. Discussion

In this study, we have identified 144 group-wisely consistent ICNs from the rsfMRI data of 77 ASD subjects and 101 healthy controls. To our best knowledge, this is one of the most comprehensive group-wise studies of ICNs among ASD studies. Due to the limitations on computational method and analysis power, previous studies usually focused on several functional networks or ICNs (Philip et al., 2012; Stigler et al., 2011). Thanks to the novel and powerful large-scale fMRI data mining framework proposed in this study and the availability of big rsfMRI dataset, it is surprising that the number of common ICNs in *the* human brain could be as large as 144 (http://hafni.cs.uga.edu/autism/templates/all.html), which is significantly larger than the number of networks analyzed in the previous studies (van den Heuvel and Hulshoff Pol, 2010; Rosazza and Minati, 2011). Moreover, our further investigations on these 144 common ICNs between ASD patients and healthy controls identified interesting patterns that may unveil atypical behaviors in ASD.

One major fMRI findings among ASD studies is the atypical patterns of Fusiform Gyrus activation during face processing in ASD (Stigler et al., 2011). These findings are directly related to the explanations of abnormal social behaviors in ASD. In our analysis, decreased connectivity has also been identified in Fusiform Gyrus during resting-state (network 144, 57, 69). These findings not only add *support* to the existing studies, but also suggest that the abnormities in task related functional networks can also be identified during resting-state.

Since language and communication impairment is one of the major symptoms in ASD domain, abnormity in brain activity during language tasks is of great research interest. For instance, decreased left IFG and increased Planum Temporale (PT) activations in subjects with ASD have been reported in several studies (Stigler et al., 2011). Intriguingly, in our analysis, we found increased interactions between DMN and language related networks such as network 84 (PSG / Right AG / Right FP), network 129 (Right IFG), and network 111 (Right MTG / Right AG / SG). Further investigations on those networks will improve our understandings *of* the language deficits in ASD.

In previous works, decreased ACG activation was recorded in ASD patients during response inhibition tasks (Kana et al., 2007) and increased rostral ACG activation was reported during correct trials of a response monitoring task (Thakkar et al., 2008). Atypical inhibition network (ACG / Middle Cingulate Gyrus (MCG) / Insular) connectivity in ASD has been hypothesized to contribute to the core repetitive behavior observed in ASD. In our analysis, the reduced ORS in network 30 (ACG / Insular) agrees with these previous findings that there is decreased connectivity within inhibition networks during resting state.

The reduced ORS in the DMN (network 7) is consistent with previous findings that ASD patients have reduced DMN connectivity (Assaf et al., 2010; Kennedy and Courchesne, 2008). *The* DMN is known to be active during resting-state and also correlate with self-referential mental representation and theory of mind (Buckner et al., 2008). Growing evidence showed correlations between the core symptom domains of ASD and DMN *involvement*. Our finding further confirmed the role of *the* DMN in the *pathophysiology* of ASD.

Intriguingly, we also found increased connections within *the* PC (network 45) which is part of *the* DMN. Although the ORS between network 45 and network 7 is relatively high (0.24), two ICNs occur in different part of PC. Network 7 covers the ventral part of PC which is adjacent to PCG while network 45 covers the dorsal part of PC which is at the top of network 7. It has been shown that *the* dorsal PC has very different roles in comparison with *the* ventral PC and is involved in spatially guided behavior and mental imagery (Zhang and Li, 2012). Together with *the* dorsal PC, increased connections were identified in *the* FP (network 137, 123), which *compels further investigation* (Okuda et al., 2003). All these observations suggest that ASD patients may have more imaginary thoughts than normal controls and that these thoughts might be relate to their deficit social behavior.

Given the large number of common ICNs identified in this study and the complicated interactions between them, it is not possible to fully investigate the roles of the ICNs in ASD in this paper at current stage. From our perspective, more participants and wider collaborations are possible solutions in fully understanding the brain functional *mechanisms* from these 144 common ICNs. To enable explorations and investigations of the results generated from the framework, we built a data portal (http://hafni.cs.uga.edu/autism/) for the dataset and the obtained ICNs. In the portal, the spatial pattern of *the* 144 common ICNs and their ontology, together with our analysis, *are* provided for further exploration and interpretation.

In addition to the novel insights on ASD studies, this work also provides a novel and effective solution for big data fMRI studies. Given the computational efficiency of Apache Spark using *connectivity map*, the method can be easily scaled up to analyze and cluster brain networks in big datasets consisting of thousands of subjects by its integration with large-scale informatics *systems* such as HAFNI-Enabled Large-scale Platform for Neuroimaging Informatics (HELPNI) (Makkie et al., 2015). In the future, the researchers can apply our proposed framework with *connectivity map* together with functional brain network decomposition methods such as SR (Lv et al., 2015a,b) to analyze such large scale brain functional datasets and identify abnormal as well as common functional networks in diseased brains *in order to better understand the* functional *mechanisms* of brain diseases.

References

Anon., 2014. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. MMWR Surveill. Summ. 63, 1–21.

Asperger, H., 1944. Die autistischen psychopathen in kindersalter. Eur. Arch. Psychiatry Clin. 1, 76–136.

Assaf, M., Jagannathan, K., Calhoun, V.D., Miller, L., Stevens, M.C., Sahl, R., O'Boyle, J.G., Schultz, R.T., Pearlson, G.D., 2010. Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. NeuroImage 53, 247–256.

Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans. Med. Imaging 23, 137–152.

Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network: anatomy, function, and relevance to disease. Ann. N. Y. Acad. Sci. 1124, 1–38.

Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. 10, 186–198 (http://www.ncbi.nlm.nih.gov/pubmed/19190637).

Chen, H., Li, K., Zhu, D., Jiang, X., Yuan, Y., Lv, P., Zhang, T., Guo, L., Shen, D., Liu, T., 2013a. Inferring group-wise consistent multimodal brain networks via multi-view spectral clustering. IEEE Trans. Med. Imaging 32, 1576–1586.

Chen, H., Zhang, T., Liu, T., 2013b. Identifying Group-Wise Consistent White Matter Landmarks via Novel Fiber Shape Descriptor. MICCAI, LNCS 8149, pp. 66–73.

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., 2006. A core system for the implementation of task sets. Neuron 50, 799–812.

Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cogn. Sci. 14, 172–179.

Fedorenko, E., Duncan, J., Kanwisher, N., 2013. Broad domain generality in focal regions of frontal and parietal cortex. Proc. Natl. Acad. Sci. U. S. A. 110, 16616–16621.

Fox, M.D., Raichle, M.E., 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat. Rev. Neurosci. 8, 700–711 (http://www.ncbi.nlm.nih.gov/pubmed/17704812).

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc. Natl. Acad. Sci. U. S. A. 102, 9673–9678.

Friston, K.J., 2009. Modalities, modes, and models in functional neuroimaging. Science 326, 399–403.

Heeger, D.J., Ress, D., 2002. What does fMRI tell us about neuronal activity? Nat. Rev. Neurosci. 3, 142–151.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17, 825–841 (http://www.ncbi.nlm.nih.gov/pubmed/12377157).

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. FSL. NeuroImage 62, 782–790.

Kana, R.K., Keller, T.A., Minshew, N.J., Just, M.A., 2007. Inhibitory control in high-functioning autism: decreased activation and underconnectivity in inhibition networks. Biol. Psychiatry 62, 198–206.

Kana, R.K., Uddin, L.Q., Kenet, T., Chugani, D., MÃ¼ller, R.-A., 2014. Brain connectivity in autism. Front. Hum. Neurosci. 8. http://dx.doi.org/10.3389/fnhum.2014.00349/abstract.

Kanner, L., 1943. Autistic disturbances of affective contact. Nerv. Child 2, 217–250.

Kennedy, D.P., Courchesne, E., 2008. Functional abnormalities of the default network during self- and other-reflection in autism. Soc. Cogn. Affect. Neurosci. 3, 177–190.

Li, K., Guo, L., Nie, J., Li, G., Liu, T., 2009. Review of methods for functional brain connectivity detection using fMRI. Comput. Med. Imaging Graph. 33, 131–139.

Liu, T., 2011. A few thoughts on brain ROIs. Brain Imaging Behav. 5, 189–202.

Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. Nature 453, 869–878.

Luxburg, U., 2007. A tutorial on spectral clustering. Stat. Comput. 17, 395–416.

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., 2013. Identifying Functional Networks via Sparse Coding of Whole Brain FMRI Signals. 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE, pp. 778–781.

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., 2015a. Sparse representation of whole-brain fMRI signals for identification of functional networks. Med. Image Anal. 20, 112–134.

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., 2015b. Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function. IEEE Trans. Biomed. Eng. 62, 1120–1131.

Lv, J., Jiang, X., Li, X., Zhu, D., Zhao, S., Zhang, T., Hu, X., Han, J., Guo, L., Li, Z., Coles, C., Hu, X., Liu, T., 2015c. Assessing effects of prenatal alcohol exposure using group-wise sparse representation of fMRI data. Psychiatry Res. 233, 254–268.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2010. Online learning for matrix factorization and sparse coding. J. Mach. Learn. Res. 11, 19–60.

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

McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. Hum. Brain Mapp. 6, 160–188.

Moseley, R.L., Ypma, R.J.F., Holt, R.J., Floris, D., Chura, L.R., Spencer, M.D., Baron-Cohen, S., Suckling, J., Bullmore, E., Rubinov, M., 2015. Whole-brain functional hypoconnectivity as an endophenotype of autism in adolescents. NeuroImage Clin. 9, 140–152.

Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., Kawashima, R., Fukuda, H., Itoh, M., Yamadori, A., 2003. Thinking of the future and past: the roles of the frontal pole and the medial temporal lobes. NeuroImage 19, 1369–1380.

Pessoa, L., 2012. Beyond brain regions: network perspective of cognition-emotion interactions. Behav. Brain Sci. 35, 158–159.

Philip, R.C.M., Dauvermann, M.R., Whalley, H.C., Baynham, K., Lawrie, S.M., Stanfield, A.C., 2012. A systematic review and meta-analysis of the fMRI investigation of autism spectrum disorders. Neurosci. Biobehav. Rev. 36, 901–942.

Rosazza, C., Minati, L., 2011. Resting-state brain networks: literature review and clinical applications. Neurol. Sci. 32, 773–785.

Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. NeuroImage 52, 1059–1069 (http://www.ncbi.nlm.nih.gov/pubmed/19819337).

Starck, T., Nikkinen, J., Rahko, J., Remes, J., Hurtig, T., Haapsamo, H., Jussila, K., Kuusikko-Gauffin, S., Mattila, M.-L., Jansson-Verkasalo, E., Pauls, D.L., Ebeling, H., Moilanen, I., Tervonen, O., Kiviniemi, V.J., 2013. Resting state fMRI reveals a default mode dissociation between retrosplenial and medial prefrontal subnetworks in ASD despite motion scrubbing. Front. Hum. Neurosci. 7, 802 (http://www.ncbi.nlm.nih.gov/pubmed/24319422).

Stigler, K.A., McDonald, B.C., Anand, A., Saykin, A.J., McDougle, C.J., 2011. Structural and functional magnetic resonance imaging of autism spectrum disorders. Brain Res. 1380, 146–161.

Thakkar, K.N., Polli, F.E., Joseph, R.M., Tuch, D.S., Hadjikhani, N., Barton, J.J.S., Manoach, D.S., 2008. Response monitoring, repetitive behaviour and anterior cingulate abnormalities in autism spectrum disorders (ASD). Brain 131, 2464–2478.

van den Heuvel, M.P., Hulshoff Pol, H.E., 2010. Exploring the brain network: a review on resting-state fMRI functional connectivity. Eur. Neuropsychopharmacol. 20, 519–534.

Van Dijk, K.R.A., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L., 2010. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J. Neurophysiol. 103, 297–321 (http://www.ncbi.nlm.nih.gov/pubmed/19889849).

Zhang, S., Li, C.R., 2012. Functional connectivity mapping of the human precuneus by resting state fMRI. NeuroImage 59, 3548–3562.

Zhu, D., Li, K., Faraco, C.C., Deng, F., Zhang, D., Guo, L., Miller, L.S., Liu, T., 2012. Optimization of functional brain ROIs via maximization of consistency of structural connectivity profiles. NeuroImage 59, 1382–1393.

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.X., Duan, L., Hu, X.X., Han, J., Wang, L., Shen, D., Miller, L.S., Li, L., Liu, T., 2013. DICCCOL: dense individualized and common connectivity-based cortical landmarks. Cereb. Cortex 23, 786–800.