ORIGINAL RESEARCH

Characterizing and differentiating task-based and resting state fMRI signals via two-stage sparse representations

Shu Zhang · Xiang Li · Jinglei Lv · Xi Jiang · Lei Guo · Tianming Liu

Abstract A relatively underexplored question in fMRI is whether there are intrinsic differences in terms of signal composition patterns that can effectively characterize and differentiate task-based or resting state fMRI (tfMRI or rsfMRI) signals. In this paper, we propose a novel two-stage sparse representation framework to examine the fundamental difference between tfMRI and rsfMRI signals. Specifically, in the first stage, the whole-brain tfMRI or rsfMRI signals of each subject were composed into a big data matrix, which was then factorized into a subject-specific dictionary matrix and a weight coefficient matrix for sparse representation. In the second stage, all of the dictionary matrices from both tfMRI/rsfMRI data across multiple subjects were composed into another big data-matrix, which was further sparsely represented by a cross-subjects common dictionary and a weight matrix. This framework has been applied on the recently publicly released Human Connectome Project (HCP) fMRI data and experimental results revealed that there are distinctive and descriptive atoms in the cross-subjects common dictionary that can effectively characterize and differentiate tfMRI and rsfMRI signals, achieving 100 % classification accuracy. Moreover, our methods and results can be meaningfully interpreted, e.g., the well-known default mode network (DMN) activities can be recovered from the *very noisy and heterogeneous aggregated big-data* of tfMRI and rsfMRI signals across all subjects in HCP Q1 release.

Introduction

Functional magnetic resonance imaging (fMRI) based on blood-oxygen-level dependent (BOLD) techniques has been widely used to study the functional activities and cognitive behaviors of the brain based on the induced stimulus by tasks, i.e., task fMRI (tfMRI) (Worsley and Friston 1995; Worsley 1997; Linden et al. 1999; Heeger and Ress 2002; Calhoun et al. 2011) or during task-free resting-state, i.e., resting state fMRI (rsfMRI) (Raichle et al. 2001; Fox and Raichle 2007). To infer meaningful neuroscientific patterns within fMRI data, various computational/statistical methods have been proposed, including the widely-used general linear model (GLM) for tfMRI (Friston et al. 1994; Worsley 1997), independent component analysis (ICA) for rsfMRI (McKeown et al. 1998), as well as many other methods, including wavelet algorithms (Bullmore et al. 2003; Shimizu et al. 2004), Markov random field (MRF) models (Descombes et al. 1998), mixture models (Hartvig and Jensen 2000), autoregressive spatial models (Woolrich et al. 2014), and Bayesian approaches (Luo and Puthusserypady 2007). In these methods, GLM is one of the most widely used methods due to its effectiveness, simplicity, robustness, and wide availability (Friston et al. 1994; Worsley 1997; Lv et al. 2014a, b).

However, a relatively underexplored question in tfMRI and rsfMRI is whether there exists intrinsic, fundamental

S. Zhang · X. Li · J. Lv · X. Jiang · T. Liu (✉)
Cortical Architecture Imaging and Discovery Lab, Department of Computer Science and Bioimaging Research Center, The University of Georgia, Athens, GA, USA
e-mail: tliu@cs.uga.edu

J. Lv · L. Guo
School of Automation, Northwestern Polytechnic University, Xi'an, China

differences in signal composition patterns which can effectively characterize and differentiate these two types of fMRI signals. As task-based fMRI is widely adopted to identify brain regions that are functionally involved in a specific task performance, while resting state fMRI is used to explore the intrinsically functionally segregation or specialization of brain regions/networks (Logothetis 2008), such differences could inspire better understanding for the organization and origination of the brain cognitive functioning. Also, determining whether participants are focusing on task during task scan or being rest during resting state scan could be very crucial for the further analysis. As far as we know, there are at least three challenges in addressing the above question. Firstly, the variability of fMRI signals across brain scans and across individual subjects could be remarkable. Despite the success of using GLM-based framework in analyzing individual brain activation patterns (e.g., Worsley and Friston 1995; Bullmore et al. 1996; Woolrich et al. 2001), it has been challenging to derive consistent fMRI activation patterns across different brains and populations due to the huge variability between individuals (Brett et al. 2002; Mueller et al. 2013). Many research studies have been done to investigate the individual variability in brain imaging, and it has been shown that there are several major sources of variability (which are often mixed): 1) the variability in the structure and its corresponding functionality between individual brains, as it has been shown that the standardized parcellation of the brain still poses a major difficulty in terms of function and microanatomy (Brett et al. 2002); 2) the variability of each individual's response to the external stimulus during tfMRI scanning, as well as their variability during resting-state, which are even more significant. For instance, it has been reported that there is significant and substantial variability in the shape of responses collected across subjects, and even across multiple scans of a single subject (Aguirre et al. 1998; Barch et al. 2013; Steinmetz and Seitz 1991); 3) and consequently, the variability in the spatial distribution of the activation patterns obtained by GLM and/or functional networks inferred by network analysis could be even larger, as mentioned in the literature (McGonigle et al. 2000; Handwerker et al. 2004).

Secondly, the amount of whole-brain, voxel-wise fMRI signals from multiple subjects could be immense. For example, high-resolution tfMRI scans from the recently publicly released Human Connectome Project (HCP) has around 150,000–200,000 time series signals for one subject during a single task/resting-state scan (Barch et al. 2013). In total, for Q1 release of HCP data, there are around 10,200,000–13,600,000 time series signals for all 60 subjects of a single task. As this dataset includes 7 tasks and 1 resting state scans, the total size will grow to 81 million. The memory capacity on a server/workstation level can barely handle such a great amount of data. Also, there would be many more subjects involved if we aim to conduct a cross-population study. Therefore, eventually

we would need a scalable computational framework with the capacity of handling the big-data of fMRI signals to any available size to obtain meaningful groupwise result.

Thirdly, there are a variety of noise sources in fMRI signals. During fMRI scans, several factors including scanner instability, experiment design deficits, and effects of susceptibility of high fields may all lead to noise (Stocker et al. 2005; Hu and Norris 2004). For an individual subject, head motion, lack of attention, and other factors that are not related to the experiment design could also introduce noise (Stocker et al. 2005). There have been various studies focused on fMRI imaging quality with enormous techniques developed for the signal de-noising and artifact removal (Simmons et al. 1999; Foland and Glover 2004; Stocker et al. 2005; Friedman and Glover 2006). However, it has been rarely explored if big-data analytic strategies such as dictionary learning and sparse representation could potentially effectively deal with such a variety of noise from the entire brains of multiple subjects.

Inspired by the successes of using sparse representation in pattern recognition (Mairal et al. 2009; Kreutz-Delgado et al. 2003; Aharon et al. 2006; Lewicki and Sejnowski 2000; Wright et al. 2010) and in brain functional imaging analysis (Lee et al. 2011; Li et al. 2009, 2012; Yamashita et al. 2008; Lv et al. 2014a, b; Li et al. 2013), in this paper, we propose a novel two-stage sparse representation framework to obtain a groupwise characterization of fMRI signals obtained during various tasks (or during resting-state), which have the capability of addressing the abovementioned three challenges. Specifically, for the first challenge, *the sparse-constrained dictionary learning method has been algorithmically shown to be capable of identifying the representative components from the given fMRI dataset as the activation maps from the fMRI study usually have little overlap* (Daubechies et al. 2009). Further, proposed framework would put the representative dictionary matrix from each individual into the same space established by the common dictionary learned at the second stage, thus dealing with the inter-subject variability problem for analysis without losing individual information. For the second challenge, the two-stage framework applies a divide-and-conquer scheme by first reducing the data of each individual to its dictionary-based representation, and then aggregating the reduced data into a new input to learn the groupwise dictionary. Using the HCP Q1 dataset as an example, after the first stage, we would learn 400 dictionary atoms from 150,000 to 200,000 signals for each of the 60 subjects (Lv et al. 2014a), *while the sparsity constraint imposed on the learning process ensures that the learned dictionaries could cover the major information of the massive number of signals.* Thus, at the second stage, the input would be of a much-reduced size (400*60), and we can learn a common dictionary of all the subjects at ease, compared with the computational load of decomposing 10,200,000–13,600,000 signals. For the third challenge, as the sparse representations learned at the

first stage capture the most prominent temporal activities and their corresponding spatial organization patterns of the brain functional signal, the individual dictionaries, which serve as the input of the second stage dictionary learning, essentially have been de-noised, since in most cases noise signals are temporally inhomogeneous and spatially scattered.

The organization of this paper is as follows: in the method section we introduce our two-stage dictionary learning framework with a running example. Then the result section provides the accuracy of classification on task/resting-state fMRI data, which serves as the main verification of the proposed framework. After that, we provide the spatial/temporal characterization of three types of the common functional components obtained by the framework, which are the main new findings of our work.

Materials and methods

Overview

The computational flowchart of the proposed framework is summarized in Fig. 1, and a running example of the framework applied on the combined dataset of working memory (WM) and resting-state (RS) fMRI is illustrated in Fig. 2. In the first stage (Fig. 1a), we apply the dictionary learning method on the whole-brain tfMRI and rsfMRI signals from each subject (in both training and testing datasets) to learn dictionaries D_t (from tfMRI) and D_r (from rsfMRI) with the corresponding loading coefficients α_t and α_r, and the example results are shown in Fig. 2b. In this work, each atom in the learned dictionary along with its loading coefficient would be termed as "functional component", since it is considered as a functional basis that constitutes the whole brain activities. Then the dictionaries D_t and D_r learned at the first stage from half of the entire subjects (i.e., training dataset) would be aggregated into one single matrix $\mathbf{S}^*$ (Fig. 1b, with an example in Fig. 2c), which serves as the input for the second-stage dictionary learning to infer a new, groupwise common dictionary D^* and loading coefficient α^* (Figs. 1c and 2d). Atoms in the common dictionary and their estimated spatial maps are then termed as "common functional component", as they are inferred groupwise and constitute the functional activity variation for all subjects involved. Further, the most discriminative atoms in the common dictionary would be selected by analyzing the loading coefficients α^* as classification features (Fig. 1f, illustrated in Fig. 2e). The selected common functional components are then used to *train a support vector machine* (SVM) for the classification of the dictionaries
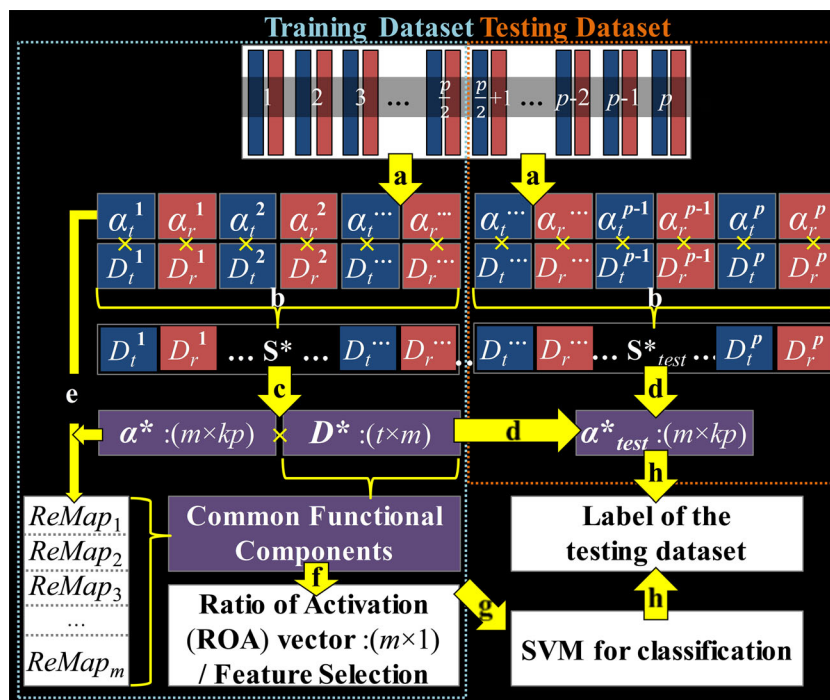
Fig. 1 Overview of the two-stage dictionary learning scheme: **a** First-stage dictionary learning routine for each individual subject and for each task type (*blue*: fMRI data obtained during task, *red*: fMRI data obtained during resting-state). α_t^1 denotes the loading coefficients obtained from the tfMRI of subject1 etc. **b** Construction of the second-stage dictionary learning input S* and sparse coding input S*$_{test}$. **c** Second-stage dictionary learning performed on S* to obtain D^* (common dictionary) and α^*. **d** Using D^* for the sparse coding on S*$_{test}$, obtaining α^*_{test}. **e** Estimation of the spatial re-maps of common functional components. **f** Calculating ROA vector by analyzing α^*, then performing classification-based feature selection. **g** Training SVM. **h** Applying SVM on α^*_{test} for the classification of the testing dataset

Fig. 2 A running example illustrating the two-stage dictionary learning framework using WM/rs fMRI dataset: **a** fMRI signals from WM (in *blue*) and resting-state (in *red*), from a total of 30 subjects; **b** Dictionaries (*upper time series plot*) and loading coefficients (*lower spatial maps*) obtained from the first stage dictionary learning. Each type of data from each subject would obtain 400 dictionaries and corresponding loading coefficients; **c** Aggregation of all learned dictionaries; **d** Common dictionaries (*left*) and their loading coefficients (*right*) obtained from the second stage dictionary learning, over a total number of 50; **e** Spatial maps of the common functional components estimated using Eq. (3). The color-coded ROA vector of common functional components are shown below, selected components are highlighted by yellow circles
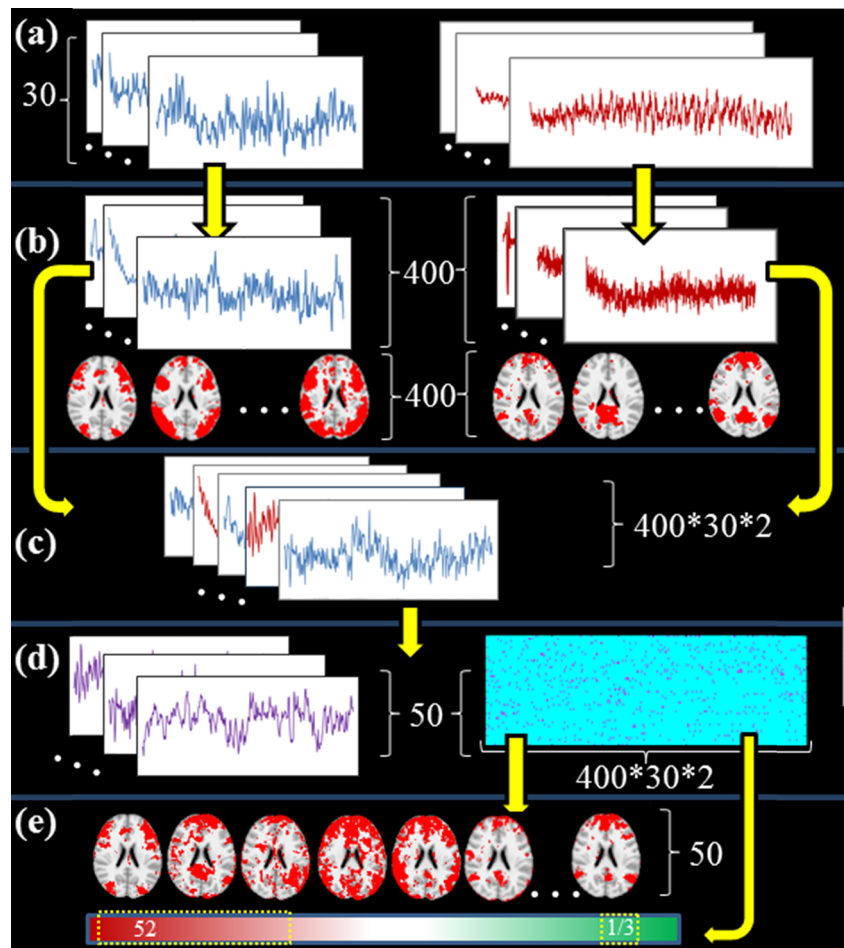
learned from the half of subjects (i.e., testing dataset) during the classification stage, as in Fig. 1g–h.

Data acquisition and preprocessing

The dataset used in this work comes from the Human Connectome Project Q1 release (Barch et al. 2013; Van Essen et al. 2013). The acquisition parameters of tfMRI data as follows: 90×104 matrix, 220 mm FOV, 72 slices, TR=0.72 s, TE=33.1 ms, flip angle=52°, BW=2290 Hz/Px, in-plane FOV=208×180 mm, 2.0 mm isotropic voxels. For tfMRI images, the preprocessing pipelines included motion correction, spatial smoothing, temporal pre-whitening, slice time correction, global drift removal. For more detailed data acquisition and preprocessing, refer to (Barch et al. 2013; Van Essen et al. 2013). rsfMRI data were acquired with the same EPI pulse sequence parameters as T-fMRI (Smith et al. 2013). The time length of each task and resting state are shown here: resting state (1200 frames), working memory (405 frames), gambling (253 frames), motor (284 frames), language (316 frames), social cognition (274 frames), relational processing (232 frames), emotion processing (176 frames). As there are 60 subjects in the released dataset, in this work half (30) of the

subjects were used for training (i.e., common dictionary learning and feature set constructing), while data from the other half were used for testing (i.e., classification). When used as dictionary learning input, signals on each voxel are normalized to have unit l_2-norm for both tfMRI and rsfMRI data.

Two-stage dictionary learning

First-stage dictionary learning method

In the first stage, the effective online dictionary learning algorithm (Mairal et al. 2009) is adopted to learn a dictionary with sparsity constraint from the whole-brain fMRI signals from *grey and white matter voxels* (with time length t and voxel number n) of each subject from both the training and testing datasets. The algorithm would learn a meaningful and over-complete dictionary D consisting of k atoms ($m>t$, $m<<n$) to represent $\mathbf{S}$ with the corresponding sparse loading coefficient matrix α, as each signal in $\mathbf{S}$ is supposed to be represented by the most relevant atoms in the learned dictionary. Specifically, for the fMRI signal set $\mathbf{S}=[s_1,s_2,\dots s_n]\epsilon\mathbb{R}^{t\times n}$, the loss function for the dictionary learning algorithm to minimize is defined in Eq. (1) with a l_1 regularization that yields to a sparse constraint

to the loading coefficient α (*constrained by non-negativity*), where λ is a regularization parameter to trade-off the regression residual and sparsity level:

$$\min_{D\epsilon\mathbb{R}^{t\times k},\alpha\epsilon\mathbb{R}^{k\times n}}\frac{1}{2}||\mathbf{S}-D\alpha||_F + \lambda||\alpha||_{1,1} \qquad (1)$$

To prevent D from arbitrarily large values which leads to trivial solution of the optimization, its columns $d_1, d_2, \ldots\ldots d_k$ are constrained by Eq. (2).

$$C \triangleq \left\{D\epsilon\mathbb{R}^{t\times k} \quad s.t. \mid \forall j = 1, \ldots k, \quad d_j^T d_j \leq 1\right\} \qquad (2)$$

In brief, dictionary learning can be rewritten as a matrix factorization problem for both D and α, and we use the effective online dictionary learning methods in (Mairal et al. 2009) to derive the solution by iteratively updating D and α in Eq. (1) during the optimization. It should be noted that we employ the same assumption as in previous studies, (Li et al. 2009, 2012; Lee et al. 2011, 2013; Oikonomou et al. 2012; Abolghasemi et al. 2013) that the atomic components (which are dictionary atoms in D in our work) involved in each voxel's fMRI signal are a few major ones, and the neural integration of those components is linear. In this work, the value of λ and dictionary size m were determined experimentally (λ=0.1, k=400) (Lv et al. 2014a, b). After the dictionary learning, the resulting D matrix contains the temporal variation of each atomic basis component of the functional brain, while the corresponding sparse loading coefficient matrix α contains the spatial distribution of each component, both illustrated in Fig. 2b.

Based on the dictionary learning results of each individual brain, our next major task is to obtain a groupwise characterization that could reveal the distinctive organization patterns between the brains' fMRI data under different conditions.

Second-stage dictionary learning and common functional components re-mapping

In this stage, all the learned dictionaries from tfMRI and rsfMRI are aggregated together to form a multi-subject, multi-type matrix $\mathbf{S}^*$ of dimension $t\times(2kp)$, where p is the number of subjects in the dataset in Fig. 2c. Note that in HCP dataset, rsfMRI data has a longer temporal length than tfMRI data for all tasks and so does the learned dictionaries. Thus we truncated the learned D_r to make them have the same length with D_t, thus enabling the aggregation of the dictionaries from different task types. $\mathbf{S}^*$ would then be used as the input for the second-stage dictionary learning analysis based on the same method as introduced previously (λ=0.1, m=50), aiming at obtaining a groupwise common dictionary D^* and the corresponding loading coefficients α^* (*constrained by non-negativity*). Compared with the original fMRI data which

are defined on the whole brain voxels of each subject, our proposed two-stage framework achieves a huge size reduction while still maintaining the major functional characterization for each individual. More importantly, noise and undesired voxel-wise signal fluctuations are largely removed in $\mathbf{S}^*$, thus we can ensure that most of the common functional components can represent the groupwise consistent functional activities, and their differences are from the intrinsic features of functional brain activity patterns. As the common dictionaries are defined on the groupwise aggregated dictionaries, it is then important to estimate their spatial maps over the brain (i.e., spatial re-mapping). In this work, the re-mapping is achieved by first aligning all the brains into the same template using linear registration. *The aligning procedure first registered the averaged frames of fMRI data into the MNI standard space of each individual subject, then the transform matrix obtained from the registration was applied to the loading coefficient matrix α of that subject, transformed it into α'. In this study, we had tried both linear and non-linear registration methods and obtained similar results for the re-mapping.* Then the spatial map of *the i-th common functional component* (*ReMap_i*) is obtained by:

$$ReMap_i = \left(\sum_{x=1}^{p}\sum_{y=1}^{k}\alpha'_{x,y_{task}}\cdot\alpha^*_{i,\,(x-1)k+y} + \sum_{x=1}^{p}\sum_{y=1}^{k}\alpha'_{x,y_{resting}}\cdot\alpha^*_{i,\,(p+x-1)k+y}\right)/2kp$$

$$(3)$$

where $\alpha'_{x,y,\,task}$ is the loading coefficient matrix of the y-th dictionary (over the total of k) of the x-th subject (over the total of p) obtained from the first stage dictionary learning on tfMRI, after registration to the template, $\alpha'_{x,y,\,resting}$ is the loading coefficient matrix of the rsfMRI result, after registration to the template, and α^*_i is the value of their corresponding loading coefficient for the i-th common dictionary from the second stage dictionary learning. In other words, the spatial maps of the common components are the weighted average from each individual component of each subject. Several sample spatial mapping results (*ReMap*) are showns in Fig. 2e.

Feature selection on common functional components

As discussed above, the common dictionaries D^* and their corresponding loading coefficients α^* obtained at the second-stage dictionary learning capture the groupwise characteristics of both types of the input fMRI data. Further, the row vectors in the loading coefficients α^* indicate the weight of the corresponding common dictionary's activation in each atom in $\mathbf{S}^*$. An example α^* matrix obtained from the WM/resting-state fMRI datasets is visualized in Fig. 2d. The $[i,j]$-th cell in α^* indicates how the i-th common dictionary is activated in the j-th atom in $\mathbf{S}^*$. As the composition of $\mathbf{S}^*$ is known in the training dataset (the pattern is illustrated in Fig. 1b: dictionaries from tfMRI and rsfMRI are put into $\mathbf{S}^*$

in turn), for the *i*-th common dictionary we can obtain its Ratio of Activation (ROA) by:

$$\text{ROA}_i = log \frac{\left|\alpha_{(i,j)}\right|_0, jth\ column\ belongs\ to\ tfMRI}{\left|\alpha_{(i,j)}\right|_0,\ jth\ column\ belongs\ to\ rsfMRI} \quad (4)$$

Thus the ratio is obtained by counting the number of non-zero entries of the row vector in **S*** which have been labeled as tfMRI or rsfMRI. A sample ROA vector for all 50 common components is visualized in Fig. 2f and color-coded by the ratio value, where a higher ratio (e.g., "4.0" in red) indicates the specific common dictionary is $e^{(4.0)}=52$ times more involved in tfMRI than in rsfMRI, while a lower value (green) indicates the opposite. ROA value approaching 1 (white) indicates that the specific component is nearly equally activated in both tfMRI and rsfMRI. Based on the ROA vector, we can then select the components that are specific to either tfMRI or rsfMRI by a high absolute value of ROA (i.e., on the two ends of the ROA vector).

In order to quantitatively define the exact set of the common functional components reflecting the underlying data composition, we design a data-driven algorithm based on the premise that the loading coefficient of the selected components shall have the maximum capacity in classifying the data. To test this premise, the algorithm would split α^* into two halves consisting of equal number of subjects (i.e., columns). Then we would use only one row from the first half of α^* which corresponds to the highest ROA value to train a Support Vector Machine (SVM) based on the LIBSVM toolbox (Chih and Chih 2011), establishing the relationship between the composition of common components (i.e., loading coefficients) and the composition of raw data (i.e., task/rs labels). Then we would use the trained SVM to classify the same rows of the second half of α^*. After storing the classification accuracy, which is defined by the proportion of columns in α^* that has been classified into the correct label, we would iteratively employ more rows in α^* sorted by their absolute ROA values as the feature inputs, thus selecting more features for the SVM

training and classification. In this way, the feature set (i.e., selected common functional components) could be determined by minimizing the classification error.

Sparse coding of the testing dataset and classification

For the purpose of verification of the proposed framework, we performed the classification analysis on the testing dataset which constitutes half of the total subjects. Before analyzing the testing dataset, the loading coefficients of the previously selected common functional components in the training dataset would be used to train an SVM in a similar way as in Feature selection on common functional components part. Note that the same first-stage dictionary learning has been performed on the testing dataset as shown in the right panel in Fig. 1a. We could aggregate the individually-learned dictionaries from the testing dataset into $\mathbf{S}^*_{testing}$, similar to the formation of $\mathbf{S}^*$ in Second-stage dictionary learning and common functional components re-mapping part. Then the common dictionary D^* obtained from the training dataset would be used to sparsely code $\mathbf{S}^*_{testing}$ by solving a typical *l*-1 regularized LASSO problem (Fig. 1d) to obtain its corresponding loading coefficients $\alpha_{testing}$:

$$\ell\left(\alpha_{testing}\right) \triangleq \min_{\alpha_{testing}\in\mathbb{R}^{m\times n}} \frac{1}{2}||\mathbf{S}_{testing}-D^*\alpha_{testing}||_F + \lambda||\alpha_{testing}||_{1,1}$$

$$(5)$$

$\alpha^*_{testing}$ has the similar implications with α^*, and the difference between them is that α^* and D^* were learned simultaneously from the training dataset utilizing an optimization routine, while $\alpha^*_{testing}$ is the deterministic LASSO solution of projecting D^* on a new dataset. As the tfMRI/rsfMRI composition pattern in $\alpha_{testing}$ is unknown, the trained SVM would be used to classify the rows in $\alpha^*_{testing}$ that correspond to feature selection results to obtain the labels of the columns in $\alpha^*_{testing}$. Thus the link between training and testing dataset is established by the fact that both of their individual dictionaries learned during the first stage are sparsely coded by the

Fig. 3 Number of components selected for the classification (x-axis) vs. component-wise classification accuracy (y-axis), using SVM-based classification method (*top panel*) and Naïve Bayesian-based classification method (*bottom panel*)
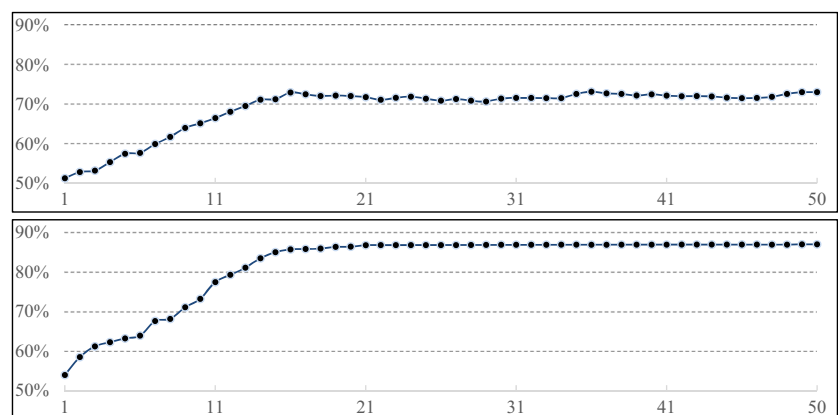
Table 1 Subject-wise classification accuracies for 7 tasks

	Emotion /rsfMRI	Gambling /rsfMRI	Languange /rsfMRI	Motor /rsfMRI	Social /rsfMRI	Relational /rsfMRI	WM /rsfMRI
#Components	24	18	20	21	23	22	16
tfMRI	100 %	100 %	100 %	100 %	100 %	100 %	100 %
rsfMRI	100 %	100 %	100 %	100 %	100 %	100 %	100 %

same common dictionary D^*, making the rows in α^* and $\alpha^*_{testing}$ corresponding to the same common functional components. After obtaining the classification result of the labels of the m number of functional components in each fMRI dataset from each subject (i.e., component-wise result), our next goal is to classify the type of that dataset (i.e., subject-wise result), as the dataset constituted by those m functional components of each subject has only one label. In this work, we used a simple scheme by comparing the number of components belonging to either task or resting-state in the given dataset, and then do the classification according to the majority voting rule.

Results

By using the HCP dataset described in Data acquisition and preprocessing, we combined each of the tfMRI data obtained from seven different tasks with one rsfMRI data, forming the seven combined datasets including emotion/rsfMRI, gambling/rsfMRI, language/rsfMRI, motor/rsfMRI, social/rsfMRI, relational/rsfMRI and working memory/rsfMRI. Then we applied the proposed framework on the seven combined datasets. In all the datasets, tfMRI and rsfMRI can be effectively differentiated, and the intrinsic spatial/temporal pattern underlying such difference could be characterized by the learned common functional components. In this work, we categorized the functional components into three types: task-evoked components, high-frequency components, and resting-state components. In most of the following sections, we would use the combined working memory (WM) tfMRI/rsfMRI dataset as an example to showcase our results, while the results from the other six tasks can be found in the supplemental materials.

Table 2 Classification accuracies on WM task / resting-state fMRI data using various dictionary size and λ values for the second stage dictionary learning

λ value	0.5	0.8	0.1	0.15	0.25	0.35
Accuracy for classification of tfMRI	100 %	100 %	100 %	97 %	97 %	87 %
Accuracy for classification of rsfMRI	93 %	90 %	100 %	87 %	43 %	77 %

Classification results on testing dataset and feature selection

As described in "Feature selection on common functional components" section, we used classification accuracy on half of the training data as the criteria for determining the exact portion of common functional components that would be used for the classification on the testing dataset. The component-wise accuracy plot obtained from WM/rsfMRI data using different numbers of features (i.e., components) *and two different classification methods (SVM and Naïve Bayesian)* is shown in Fig. 3. It can be seen that when the number of features used was small, the classification performance is only slightly better than random guess. As more components were used, the accuracy increased monotonically and then reached the maximum at 16 *for both classification methods*. As the performance would not change much afterwards, we could conclude that the additional components employed did not contribute much to the differentiation power, thus totally 16 components were selected as the features for classification.

After the feature selection in each of the seven combined task/rsfMRI datasets, we classified their corresponding testing datasets following steps in Sparse coding of the testing dataset and classification part, and the subject-wise results are summarized in Table 1. It can be seen that the classification accuracies are very high: tfMRI data from all the 30 subjects have been classified correctly, rsfMRI data from all the 30 subjects also have been classified correctly *using both SVM-based and Naïve Bayesian-based classification methods*. The results demonstrate that there exists fundamental differences between the component composition of tfMRI and rsfMRI, while the common functional components (i.e., features for the classification input) learned by the proposed model has the capability for uncovering and characterizing such differences from the large and noisy groupwise data.

In Table 1, the first row shows the number of common functional components used for the classification by feature selection. The second row shows the percentage of tfMRI dataset of all 30 subjects that has been classified to the correct label. Similarly, the third row shows the percentage of rsfMRI dataset classified to the correct label. *To further investigate the effect of the regularization parameter λ value on the classification results, we have tested the framework on the same WM/rsfMRI dataset with various λ values, the final classification accuracies are shown in*

Table 3 Subject-wise classification accuracies for 7 tasks using reduced dictionary size of 25 for second stage dictionary learning

Dictionary size=25	Emotion /rsfMRI	Gambling /rsfMRI	Languange /rsfMRI	Motor /rsfMRI	Social /rsfMRI	Relational /rsfMRI	WM /rsfMRI
tfMRI	100 %	100 %	93 %	100 %	100 %	100 %	100 %
rsfMRI	87 %	70 %	57 %	93 %	100 %	57 %	83 %

Table 2. The results show that the classification accuracy would be relatively stable within a stable range, especially for the performance on tfMRI dataset. However, extreme larger λ value would lead to a loading coefficient matrix (i.e., input feature for classification) that is too sparse, which decreases the differentiation capability of the features and reduces the classification accuracy. Also, the classification accuracies of 7 task/rsfMRI datasets using reduced dictionary size of 25 in the second stage dictionary learning are listed in Table 3. The results shown that although the tfMRI dataset could be identified accurately using smaller dictionary size (and consequently less number of features to use), dataset from rsfMRI could not be successfully distinguished from certain tasks, indicating the importance of the framework to effectively cover the whole component space by using a sufficiently large common dictionary size during the learning.

Task-evoked common functional components

The most prominent and intuitive common functional components obtained by our framework is the task-evoked type. In the working memory task, there is an example component that belongs to this category, with very high ROA values of 4.1 and it has been selected for the classification. The spatial distributions of this component is very similar to the results from groupwise GLM activation detection applied on the tfMRI of WM task from the 30 subjects in training dataset, as shown Fig. 4a and b, where the spatial overlapping rate between (a) and (b) are 89.5 %. Its time series, plotted in

Fig. 4 Example task-evoked common functional component from WM/RS dataset. **a**: volume map of the component; **b** volume map of the corresponding contrast map by groupwise GLM; **c**: component mapped on inflated cortical surface; **d** groupwise GLM result mapped on inflated cortical surface; **e** time series of the components (*blue*), task design contrast curve of WM task (*yellow*); **f**: frequency spectrum of the components (*red*), frequency spectrum of the contrast curve (*green*)
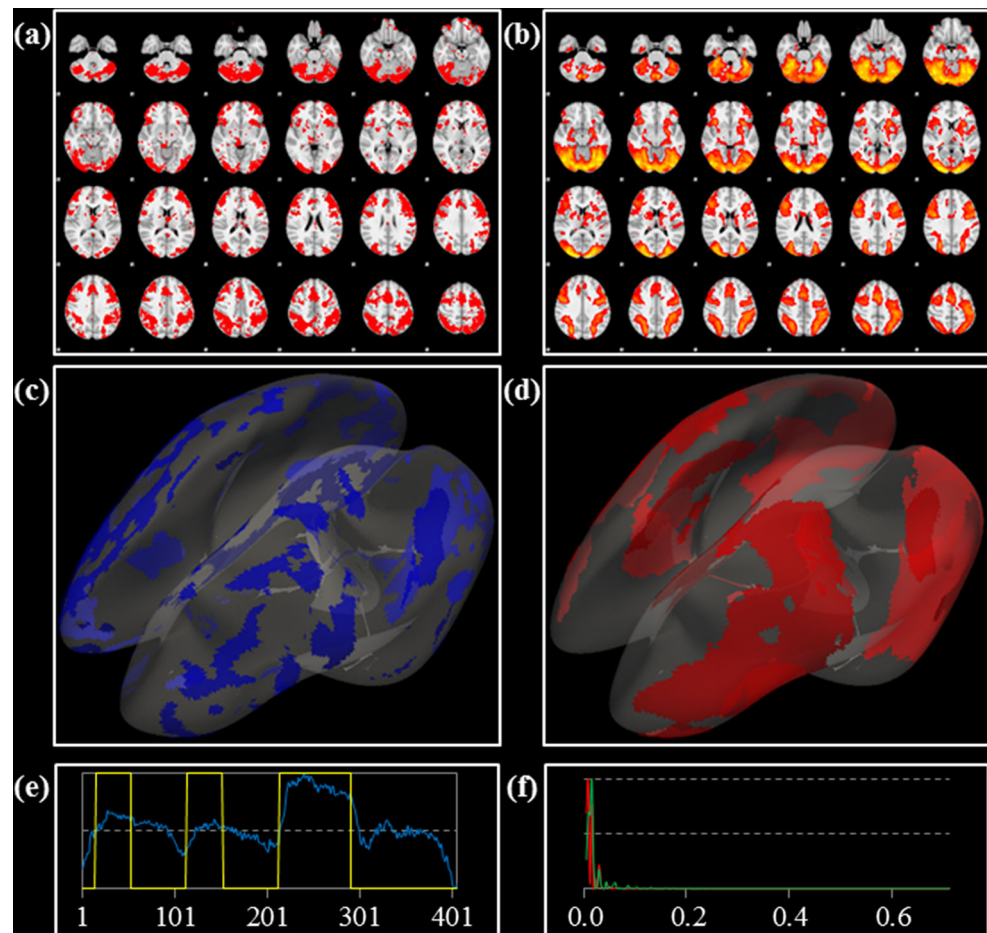
Fig. 4e, are correspondent with the task design contrast curves (correlation value: 0.6653). Further, the frequency spectrum of its time series (Fig. 4f) is highly concentrated on the task design frequency. Based on the spatial, temporal, frequency-domain characteristic and its sole presence in tfMRI, we can be assured that our framework could identify task-evoked functional component in the large scale combined fMRI data. More results could be found in supplemental materials (Supplemental Figs. 1–6).

Resting-state domain common functional components

Opposite to the task-evoked components, there is one resting-domain common functional component with the lowest ROA value=−1.1 (i.e., the most frequently activated in rsfMRI) in the WM/RS dataset. As visualized in Fig. 5a, its spatial map largely resembles the widely-reported default mode network (DMN) (Raichle et al. 2001). We had also applied the groupwise independent component analysis (ICA) on the same dataset and obtained similar pattern, as shown in Fig. 5b. It should be noted that as no low-pass filtering has been applied in HCP rsfMRI pre-processing, the dominance of lower frequency in the component spectrum (Fig. 5f) is a

valid characterization of the resting-state brain functional activation pattern, rather than from the filtering artifact (spatial overlapping rate with ICA resting-state map: 83 %). More results could be found in supplemental materials (Supplemental Figs. 7–12).

High frequency common functional components

Besides the two traditional types of common functional components described above, several of the identified components from various tasks are immensely activated in tfMRI data, yet exhibit diverse spatial/temporal patterns, compared with the common knowledge of brain regions that are responding to tasks. One characteristic shared by those components is the dominance of high frequency in their spectrum (bottom panel of Fig. 6). It is interesting that components from various datasets have almost the same frequency domain characteristics and very similar spatial distribution, even though the task design and time length are all different in these datasets. By examining the spatial map of those components in Fig. 6, it could be found that in all the three tasks (WM, emotion and gambling) the ventral posterior cingulate cortex is consistently activated, which receives inputs from thalamus and neocortex,

Fig. 5 Example resting-state common functional component from WM/RS dataset. a: volume map of the component; b volume map of the corresponding groupwise ICA result; c: component mapped on inflated cortical surface; d groupwise ICA result mapped on inflated cortical surface; e time series of the component; f: frequency spectrum of the component
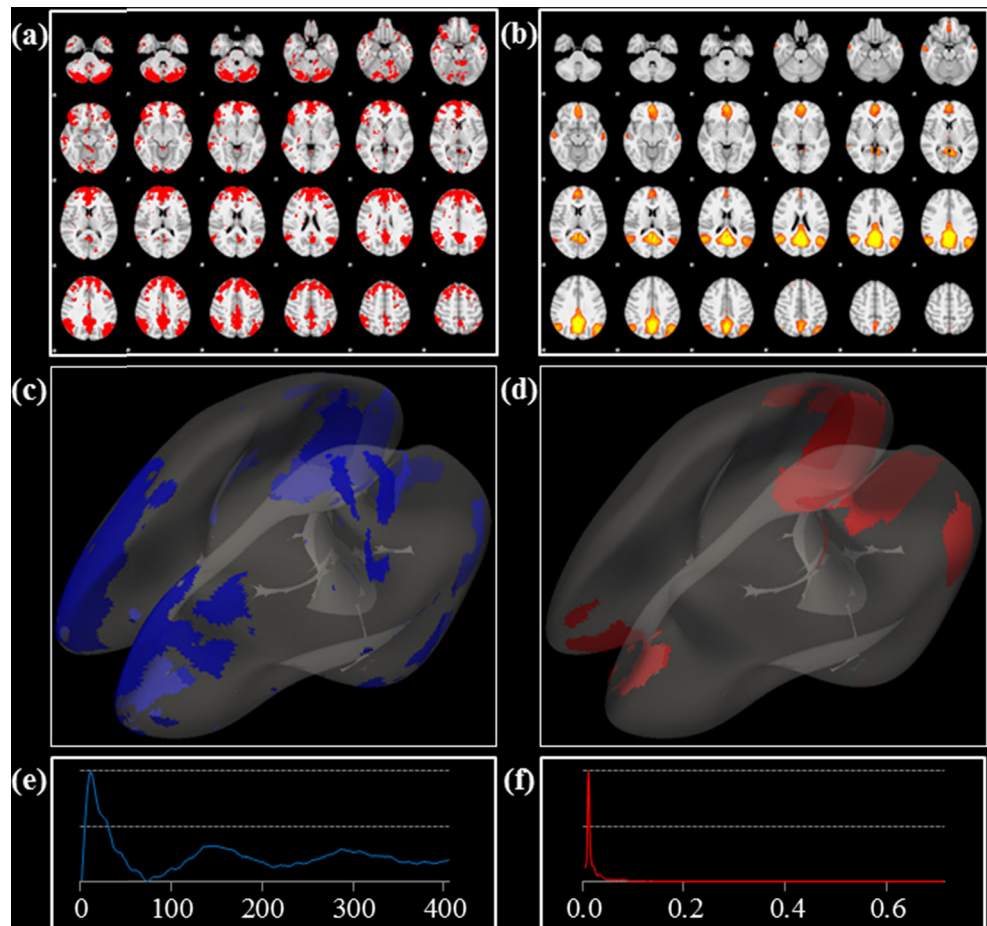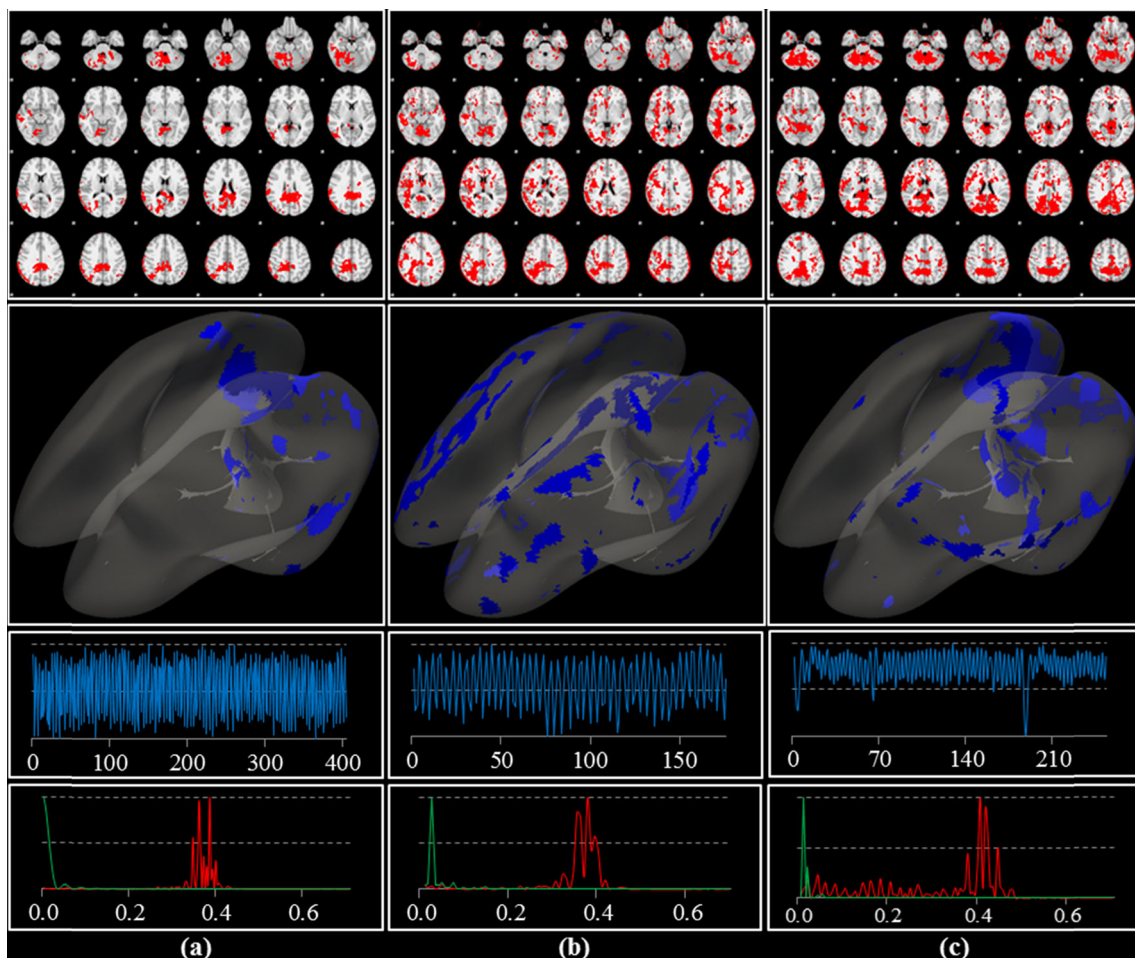
Fig. 6 High frequency common functional components identified from three datasets: **a** WM/RS, **b** Emotion/RS, and **c** Gambling/RS. First (*top*) panel: volume maps of the components; second panel: component mapped on inflated cortical surface; third panel: time series of the components; fourth (*bottom*) panel: frequency spectrums of the components (*red*), frequency spectrums of the contrast curves are shown in *green*

and projects to the entorhinal cortex via cingulum. Being an integral part of the limbic system, this area has been reported to be involved with associative learning (Maddock et al. 2001), memory retrieval (Nielsen et al. 2005), as well as emotion formation and processing (Maddock et al. 2003), which explains its significant presence during those tasks. Unlike resting-state networks which have been reported to be at presence in different tasks with similar spatial distribution (e.g., DMN) (Raichle et al. 2001), the common functional components shown in Fig. 6 only activate during their respective tasks but rarely during resting-state, thus largely excluding the possibility that these two components belong to the traditional resting-state network. Also, these components could not be identified by traditional activation detection method due to the high-frequency nature of their temporal pattern (third panel in Fig. 6), although these components only activated during the task and were highly related to tfMRI data (all with ROA value of infinity). While in our two-stage dictionary learning framework such components are very obvious and

could be robustly identified. More results could be found in supplemental materials (Supplemental Figs. 13–16).

Discussion and conclusion

By using *the HCP public* tfMRI/rsfMRI datasets, we have presented a novel two-stage sparse representation framework to examine the intrinsic differences in tfMRI/rsfMRI signals. The major methodological novelty of the two-stage sparse representation is that the framework can effectively remove the noise and undesired voxel-wise signal fluctuations, efficiently deal with the big-data (a matrix of millions times hundreds data points), and infer distinctive and descriptive common dictionary atoms that can well characterize and differentiate tfMRI/rsfMRI signals in task performance and resting state. In addition, the results also suggest that our two-stage sparse representation method can effectively recover the DMN activities from the *very noisy and heterogeneous*

Brain Imaging and Behavior

aggregated big-data of tfMRI and rsfMRI signals across all subjects in HCP Q1 release. The applications of this framework on seven HCP tfMRI datasets and one rsfMRI dataset have demonstrated promising results. In the future, we plan to better interpret other dictionary atoms in two stages and apply this framework to clinical fMRI datasets to elucidate possible alterations of functional activities in brain disorders.

Acknowledgments T Liu was supported by NSF CAREER Award (IIS-1149260), NIH R01 DA-033393, NIH R01 AG-042599, NSF CBET-1302089 and NSF BCS-1439051. *L Guo was supported by the NSFC #61273362.*

Conflict of Interest Shu Zhang, Xiang Li, Jinglei Lv, Xi Jiang, Lei Guo, and Tianming Liu declare that they have no conflicts of interest.

Informed Consent Data used in this study were previously collected and archived in a data bank.

References

Abolghasemi, V., Ferdowsi, S., Sanei, S. (2013). Fast and incoherent dictionary learning algorithms with application to fMRI. *Signal, Image and Video Processing.*

Aguirre, G. K., Zarahn, E., & D'esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *NeuroImage, 8*(4), 360–369.

Aharon, M., Elad, M., & Bruckstein, A. (2006). K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing, 54*(11), 4311–4322.

Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J.M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A.Z., Van Essen, D.C., WU-Minn HCP Consortium. (2013). Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage.*

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience, 3*(3), 243–249.

Bullmore, E., Brammer, M., Williams, S., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., & Sham, P. (1996). Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine, 35*(2), 261–277.

Bullmore, E., Fadili, J., Breakspear, M., Salvador, R., Suckling, J., & Brammer, M. (2003). Wavelets and statistical analysis of functional magnetic resonance images of the human brain. *Statistical Methods in Medical Research, 12*(5), 375–399.

Calhoun, V.D., et al. (2011). fMRI Activation in a visual-perception task: network of areas detected using the general linear model and independent components analysis. *NeuroImage, 14*(5), 1080–1088, 2001.

Chih C.C., & Chih J.L. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*:27:1–27:27.

Daubechies, I., Roussos, E., Takerkart, S., Benharrosh, M., Golden, C., D'Ardenne, K., Richter, W., Cohen, J. D., & Haxby, J. (2009). Independent component analysis for brain fMRI does not select for independence. *Proceedings of the National Academy of Sciences of the United States of America, 106*(26), 10415–10422.

Descombes, X., Kruggel, F., & von Cramon, D. Y. (1998). fMRI signal restoration using a spatio-temporal markov random field preserving transitions. *NeuroImage, 8*(4), 340–349.

Foland, L., & Glover, G.H. (2004). Scanner quality assurance for longitudinal or multicenter fMRI studies, In International Society for Magnetic Resonance Imaging. *12th Annual Meeting of the International Society for Magnetic Resonance Imaging (ISMRM).*

Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Review Neuroscience, 8*, 700–711.

Friedman, L., & Glover, G. H. (2006). Report on a multicenter fMRI quality assurance protocol. *Journal of Magnetic Resonance Imaging, 23*(6), 827–839.

Friston, KJ., Holmes, AP., Worsley, KJ. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping,* V2-I4: 189–210.

Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage, 21*(4), 1639–1651.

Hartvig, N. V., & Jensen, J. L. (2000). Spatial mixture modeling of fmri data. *Human Brain Mapping, 11*(4), 233–248.

Heeger, D. J., & Ress, D. (2002). What does fMRI tell us about neuronal activity? *Nature Review Neuroscience, 3*(2), 142–152.

Hu, X., & Norris, D. G. (2004). Advances in high-field magnetic resonance imaging. *Annual Review of Biomedical Engineering, 6*, 157–184.

Kreutz-Delgado, K., Murray, J. F., Rao, B. D., Engan, K., Lee, T. W., & Sejnowski, T. J. (2003). Dictionary learning algorithms for sparse representation. *Neural Computation, 15*(2), 349–396.

Lee, K., Tak, S., & Ye, J. C. (2011). A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion. *IEEE Transactions on Medical Imaging, 30*(5), 1076–1089.

Lee, J., Jeong, Y., Ye, J.C. (2013). Group sparse dictionary learning and inference for resting-state fMRI analysis of Alzheimer's disease. *ISBI.*

Lewicki, M., & Sejnowski, T. (2000). Learning overcomplete representations. *Neural Computation, 12*(2), 337–365.

Li, Y., Namburi, P., Yu, Z., Guan, C., Feng, J., & Gu, Z. (2009). Voxel selection in FMRI data analysis based on sparse representation. *IEEE Transactions on Biomedical Engineering, 56*(10), 2439–2451.

Li, Y., Long, J., He, L., Lu, H., Gu, Z., et al. (2012). A sparse representation-based algorithm for pattern localization in brain imaging data analysis. *PLoS ONE, 7*(12), e50332.

Li, X., Zhu, D., Jiang, X., Jin, C., Zhang, X., Guo, L., Zhang, J., Hu, X., Li, J., Liu, T. (2013). Dynamic functional connectomics signatures for characterization and differentiation of PTSD Patients, in press, *Human Brain Mapping.*

Linden, D. E., Prvulovic, D., Formisano, E., Vollinger, M., Zanella, F. E., Goebel, R., & Dierks, T. (1999). The functional neuroanatomy of target detection: an fMRI study of visual and auditory oddball tasks. *Cerebral Cortex, 9*(8), 815–823.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature, 453*(7197), 869–878.

Luo, H., & Puthusserypady, S. (2007). fMRI data analysis with nonstationary noise models: a Bayesian approach. *IEEE Transactions on Biomedical Engineering, 54*, 1621–1630.

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. (2014a). Sparse representation of whole-brain FMRI signals for identification of functional networks, in press, *Medical Image Analysis.*

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. (2014b). Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function, accepted, *IEEE Transactions on Biomedical Engineering.*

Springer

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2001). Remembering familiar people: the posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience, 104*(3), 667–676.

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human Brain Mapping, 18*(1), 30–41.

Mairal, J., Bach, Francis., Ponce, J., Sapiro, G. (2009). Online dictionary learning for sparse coding. In Proceedings of the International Conference on Machine Learning (ICML).

McGonigle, D. J., Howseman, A. M., Athwal, B. S., Friston, K. J., Frackowiak, R. S. J., & Holmes, A. P. (2000). Variability in fMRI: an examination of intersession differences. *NeuroImage, 11*(6), 708–734.

McKeown, M. J., et al. (1998). Spatially independent activity patterns in functional MRI data during the Stroop color-naming task. *PNAS, 95*(3), 803.

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T., Sepulcre, J., Sabuncu, M. R., Shafee, R., Lu, J., & Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron, 77*(3), 586–595.

Nielsen, F. A., Balslev, D., & Hansen, L. K. (2005). Mining the posterior cingulate: segregation between memory and pain components. *NeuroImage, 27*(3), 520–532.

Oikonomou, V. P., Blekas, K., & Astrakas, L. (2012). A sparse and spatially constrained generative regression model for fMRI data analysis. *IEEE Transactions on Biomedical Engineering, 59*(1), 58–67.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 676–682.

Shimizu, Y., Barth, M., Windischberger, C., Moser, E., & Thurner, S. (2004). Wavelet-based multifractal analysis of fMRI time series. *NeuroImage, 22*, 1195–1202.

Simmons, A., Moore, E., & William, S. C. R. (1999). Quality control for functional magnetic resonance imaging using automated data analysis and shewhart charting. *Magnetic Resonance in Medicine, 41*(6), 1274–1278.

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., WU-Minn HCP Consortium. (2013). Resting-state fMRI in the Human Connectome Project. *Neuroimage*.

Steinmetz, H., & Seitz, R. J. (1991). Functional anatomy of language processing: neuroimaging and the problem of individual variability. *Neuropsychologia, 29*, 1149–1161.

Stocker, T., Schnneider, F., Klein, M., Habel, U., Kellermann, T., Ziles, K., & Shah, N. J. (2005). Automated quality assurance routines for fMRI data applied to a multicenter study. *Human Brain Mapping, 25*(2), 237–246.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., & Ugurbil, K. (2013). WU-Minn HCP consortium. The WU-Minn human connectome project: an overview. *NeuroImage, 80*(2013), 62–79.

Woolrich, M. W., Ripley, B., Brady, J., & Smith, S. (2001). Temporal autocorrelation in univariate linear modelling of FMRI data. *NeuroImage, 14*(6), 1370–1386.

Woolrich, M. W., Jenkinson, M., Brady, J. M., & Smith, S. M. (2014). Fully bayesian spatio-temporal modeling of fmri data. *IEEE Transactions on Medical Imaging, 23*(2), 213–231.

Worsley, K. J. (1997). An overview and some new developments in the statistical analysis of PET and fMRI data. *Human Brain Mapping, 5*(4), 254–258.

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time series revisited again. *NeuroImage, 2*, 173–181.

Wright, J., et al. (2010). Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE, 98*(6), 1031–1044.

Yamashita, O., Sato, M. A., Yoshioka, T., Tong, F., & Kamitani, Y. (2008). Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. *NeuroImage, 42*(4), 1414–1429.