

Assessing the effects of cocaine dependence and pathological gambling using group-wise sparse representation of natural stimulus fMRI data

Yudan Ren¹ · Jun Fang¹ · Jinglei Lv^{1,2} · Xintao Hu¹ · Cong Christine Guo³ · Lei Guo¹ · Jiansong Xu⁴ · Marc N. Potenza⁴ · Tianming Liu²

Published online: 4 October 2016
© Springer Science+Business Media New York 2016

Abstract Assessing functional brain activation patterns in neuropsychiatric disorders such as cocaine dependence (CD) or pathological gambling (PG) under naturalistic stimuli has received rising interest in recent years. In this paper, we propose and apply a novel group-wise sparse representation framework to assess differences in neural responses to naturalistic stimuli across multiple groups of participants (healthy control, cocaine dependence, pathological gambling). Specifically, natural stimulus fMRI (N-fMRI) signals from all three groups of subjects are aggregated into a big data matrix, which is then decomposed into a common signal basis dictionary and associated weight coefficient matrices via an effective online dictionary learning and sparse coding method. The coefficient matrices associated with each common dictionary atom are statistically assessed for each group separately. With the inter-group comparisons based on the group-wise correspondence established by the common dictionary, our experimental results demonstrated that the group-wise sparse coding

and representation strategy can effectively and specifically detect brain networks/regions affected by different pathological conditions of the brain under naturalistic stimuli.

Keywords Functional brain networks · Naturalistic stimuli · Group-wise sparse representation · Cocaine dependence · Pathological gambling

Introduction

In recent years, assessing group differences of functional brain activity by functional resonance imaging (fMRI) has drawn increasing attention (Lv et al. 2015c; Gur et al. 2000; Baron-Cohen et al. 1999; Manoach et al. 2000; Tapert et al. 2001; Codispoti et al. 2008; Kober et al. 2016; Kret and De Gelder 2012). However, most previous studies were based on task fMRI with abstracted and repeated stimuli (Lv et al. 2015a, b, c; Gur et al. 2000; Baron-Cohen et al. 1999; Manoach et al. 2000; Tapert et al. 2001), while only a handful of studies employed natural stimuli fMRI (N-fMRI) such as video watching (Codispoti et al. 2008; Kober et al. 2016; Kret and De Gelder 2012). Compared to task fMRI, N-fMRI is more complex, dynamic and closer to human brain's daily perception, which provides multiple cognitive loads in an uncontrolled environment and thus alternative, new opportunities to better understand the brain's functional activities (Hasson et al. 2010). One of the underlying difficulties of N-fMRI analysis is that it's hard to build correspondence between the input naturalistic stimuli and any specific brain function (Bordier et al. 2013). Thus it is challenging to use popular model-driven methods like general linear model (GLM) to analyze N-fMRI data (Bordier et al. 2013). Although there are already a few studies applying GLM on N-fMRI data, prior manual rating processes are typically required for each

Electronic supplementary material The online version of this article (doi:10.1007/s11682-016-9596-4) contains supplementary material, which is available to authorized users.

✉ Marc N. Potenza
marc.potenza@yale.edu

✉ Tianming Liu
tliu@cs.uga.edu

¹ School of Automation, Northwestern Polytechnical University, Xi'an, China

² Cortical Architecture Imaging and Discovery Lab, Department of Computer Science and Bioimaging Research Center, The University of Georgia, Athens, GA, USA

³ QIMR Berghofer Medical Research Institute, Brisbane, Australia

⁴ Department of Psychiatry, Yale University, New Haven, CT, USA

feature of video stimuli, which is a complex, time-consuming procedure and might lead into inter-rater difference and/or artefacts (Lahnakoski et al. 2012). Meanwhile, data-driven methods such as independent component analysis (ICA) has been another school of widely used methodologies in exploring functional brain networks in N-fMRI, which aims to decompose fMRI signals into meaningful components and does not need any priori hypothesis of possible incentives of brain responses (Bartels and Zeki 2004).

Recently, there are growing numbers of research studies that decompose fMRI data into linear combinations of multiple components via sparse representation of whole-brain fMRI signals (Lv et al. 2015a; Lv et al. 2015b). The basic idea of this framework is to extract whole-brain fMRI signals of one subject and aggregate them together into a big data matrix, which is then decomposed into an over-complete dictionary matrix and a weight coefficient matrix by effective dictionary learning and sparse coding algorithms (Mairal et al. 2010). In this framework, each dictionary atom stands for the representative signal pattern and functional activities of a brain network and its associated weight coefficient vector stands for the spatial distribution of this dictionary atom. One interesting characteristic of this framework is that the decomposed weight coefficient matrix naturally indicates the spatial overlap/interaction pattern among reconstructed brain networks. This novel data-driven method naturally considers that a brain region might be involved in multiple functional processes/domains (Gazzaniga 2004; Pessoa 2012), and therefore, each fMRI signal is factorized into several network atoms (Lv et al. 2015a; Lv et al. 2015b).

However, one notable challenge of previous studies of sparse representation of whole-brain fMRI signals is how to establish the correspondence of different dictionary atoms across individual subjects (Lv et al. 2015c). It would be difficult to assess group differences of functional brain activities in different brain conditions such as CD, PG and healthy controls in this paper without established group-wise correspondence of such dictionary atoms. Responding to this challenge, in this paper, we propose a novel computational framework of group-wise sparse coding and representation of the N-fMRI data of multiple groups of subjects (healthy control, CD, PG in this work) and assess the group differences of functional brain networks. Comparing with previous work that applied group-wise sparse representation approach to task fMRI data (Lv et al. 2015c), we employed and expanded this effective framework to N-fMRI data with more complex and dynamic stimuli from multiple individuals, which is the major technical novelty of our study. Although N-fMRI involves dynamic multimodal stimuli with rich contexts, it has been reported that different individuals tend to respond in similar way under complex naturalistic stimuli (Hu et al. 2015; Han et al. 2015). In addition, the multiple cognitive loads in naturalistic stimuli provide us the opportunities to have better understanding of brain activation patterns in different brain conditions.

Therefore, by applying our effective group-wise sparse representation method to N-fMRI can not only identify the functional brain networks of interest, but also can uncover group-wise differences in brain activation patterns among different conditions. More specifically, our computational framework aims to learn a common time series dictionary matrix from the aggregated fMRI signals of all three groups of subjects, and subsequently assess weight coefficient matrices corresponding to each common dictionary atom for each group separately. With comparisons of the inter-group differences based on the correspondence established by common dictionaries, our experimental results demonstrated that the group-wise sparse coding strategy can effectively reveal different brain responses of CD, PG and healthy controls under different naturalistic stimuli in a collection of brain networks/regions.

Methods

Overview

The overview of our framework is summarized in Fig. 1. First, subjects from 3 groups (HC: healthy control, CD: cocaine dependence, PG: pathological gambling) are spatially normalized into the standard MNI template. Then a standardized group common brain mask is used to extract the whole brain fMRI signals of each subject (Lv et al. 2015c), and the extracted signals are stacked into 2D signal matrix $S_x \in \mathbb{R}^{f \times n_x}$, as shown in Fig. 1(a). Then all extracted signal matrices form 3 groups of subjects are pooled into a big matrix $S \in \mathbb{R}^{f \times n}$, which is composed of three groups of subjects: G_{CD} , G_{HC} and G_{PG} (Fig. 1(b)). The online dictionary learning and sparse coding is adopted (Mairal et al. 2010), and the signal matrix S is decomposed into a common signal dictionary matrix D and associated coefficient matrix A (Fig. 1(c)). Specifically, D is commonly shared by three groups, and A has the same spatial voxel organization and group correspondence as S , i.e., $A = [A_{G_{CD}}, A_{G_{HC}}, A_{G_{PG}}] \in \mathbb{R}^{m \times n}$ (Fig. 1(c)). The group-wise correspondence established by the common dictionary D provides the opportunity for group-wise statistical analysis and comparison (Fig. 1(d)). To compare brain activation among three groups, a statistical coefficient mapping method is applied to sub-coefficient matrix of each group to derive statistical coefficient maps (z-score maps) as shown in Fig. 1(e) (Lv et al. 2015c). Then, the z-score maps are employed for group comparison analysis to identify brain regions/networks related to different brain conditions (Fig. 1(d)).

Data acquisition and pre-processing

Video stimuli Three videos are selected including cocaine, gambling and sad scenarios presented in a counter-balanced

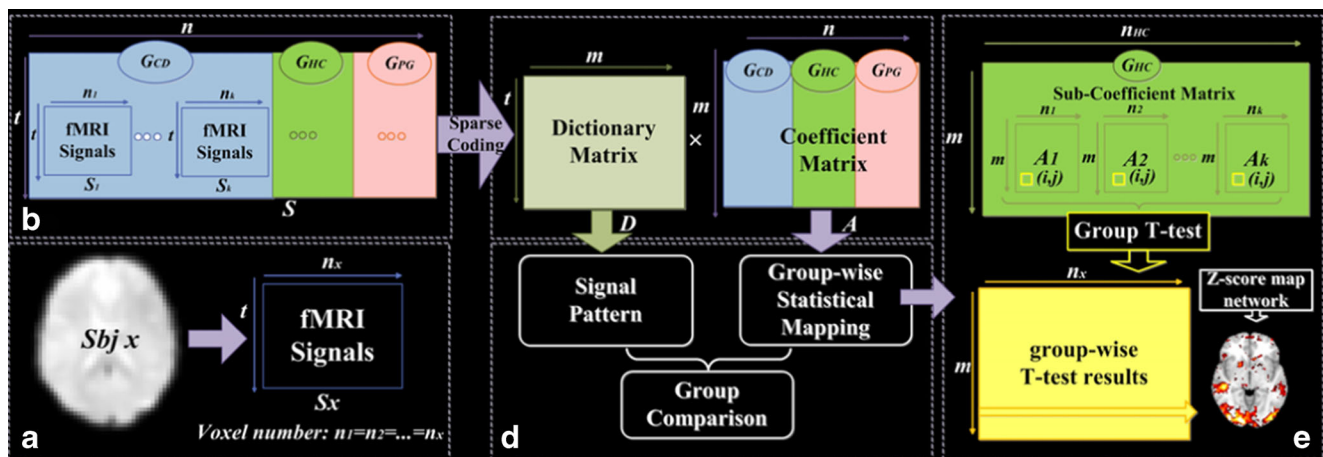


Fig. 1 The computational framework of our group-wise sparse representation of whole-brain fMRI signals from three different groups of subjects (CD: Cocaine dependence, HC: Healthy control, PG: Pathological gambling)

order. In each video, a female actor was shown speaking to the participants and describing a cocaine-use, gambling, or sad experience (Kober et al. 2016). The lengths of cocaine, gambling and sad experience videos are 198, 217, and 218 s respectively. There was a 30–45 s of baseline (gray screen) before and after each tape.

Participants Forty-four participants (20 females) were included in this study, of whom 14 were CD (7 female), 15 were PG (5 female), and 15 HC were neither (8 female). Participants were inspected by phone-screening to determine initial eligibility and were excluded if they were left-handed, did not speak English, were treatment-seeking, or reported head trauma with loss of consciousness, pregnancy, claustrophobia, or any implants or non-removable metal contraindicated in MRI. Eligible participants were administered a structured clinical interview for DSM-IV (SCID) and participants who met criteria for an active (past-3-month) axis I disorder (except past/present CD, PG, or nicotine dependence), were reported a history of neurological illness, psychiatric hospitalization in the last 6 months (except for CD), or use of psychotropic medications were excluded from the study. Written informed consent was obtained from all participants after complete description of the study. The study was approved by the Yale Human Investigation Committee. This study employed two widely-used, valid and reliable measures, the South Oaks Gambling Screen (Lesieur and Blume 1987) and Fagerstrom Test for Nicotine Dependence (Heatherton et al. 1991) to assess gambling and smoking, respectively.

Imaging data acquisition and preprocessing

Participants were scanned in a 3 T Siemens Trio MRI system. Functional images were acquired in the axial plane parallel to the AC-PC line (TR/TE = 1500/27 ms, flip angle = 60°, field of view = 220x220mm, and 25x5mm slices). Each

type of video was presented to all three groups of subjects in scanner with same sequence, and the time points of functional images corresponding to cocaine-use, gambling, and sad experience videos were 198, 217, and 218. In addition, high-resolution 3D Magnetization Prepared Rapid Gradient Echo structural images were acquired for multi-subject registration (TR/TE = 2530/3.34 ms; flip angle = 7°, field of view = 256 × 256 mm, and 176x1mm slices). The preprocessing pipeline is implemented using Data Preprocessing Assistant for Resting-state FMRI (DPARSF, <http://rfmri.org/DPARSF>). The preprocessing pipeline includes motion correction, slice time correction, spatial smoothing, band-filtering (0.008–0.3), and normalization. We generated a group-wise common mask by conducting all single masks together and use this common mask to extract whole brain signals of each subject (Lv et al. 2015c).

Dictionary learning and sparse representation of whole-brain fMRI signals

As many previous studies have reported, a variety of brain regions and networks exhibit strong functional diversity and heterogeneity, which means that a cortical region could involve in multiple functional processes and a functional networks might contain various heterogeneous neuro anatomic regions (Gazzaniga 2004; Pessoa 2012). The sparse representation of whole-brain fMRI signals can effectively and robustly identify concurrent functional networks, not only for task-evoked networks but also for resting-state networks (Lv et al. 2015a; Lv et al. 2015b). Despite the successful applications of sparse representation of whole-brain fMRI signals, conducting sparse representation method to whole-brain naturalistic stimulus fMRI of brain conditions has rarely been explored yet, as far as we know.

Applying dictionary learning and sparse coding method to a big signal set from whole brains of all the subjects aims at

minimizing the representation error and learning an efficient dictionary to represent each signal by a sparse set of relevant dictionary components. The empirical cost function is summarized in Eq. (1) where the average loss of n signals is taken into consideration. The loss function of each signal is defined as Eq. (2). The L1 regularization is adopted for sparsity control.

$$f_n(D) \triangleq \frac{1}{n} \sum_{i=1}^n l(s_i, D), \quad (1)$$

$$l(s_i, D) \triangleq \min_{a_i \in \mathbb{R}^m} \frac{1}{2} \|s_i - DA_i\|_2^2 + \lambda \|A_i\|_1 \quad (2)$$

Notably, each s_i in S is normalized with zero mean and standard deviation of 1 to ensure the coefficients in A comparable. Also, each column in D is constrained with Eq. (3).

$$C \triangleq \left\{ D \in \mathbb{R}^{t \times m} \text{ s.t. } \forall j = 1, \dots, m, \quad d_j^T d_j \leq 1 \right\} \\ \min_{D \in C, a \in \mathbb{R}^{m \times n}} \frac{1}{2} \|S - DA\|_F^2 + \lambda \|A\|_{1,1} \quad (3)$$

In short, dictionary learning and sparse coding can be summarized as Eq. (3) (Mairal et al. 2010). The online dictionary learning method provides us a platform to learn the dictionary and representation efficiently and optimally. In our studies, each group of subjects (CD, HC, PG) viewed the same three videos (cocaine-use, gambling, and sad experience), thus we conduct group-wise dictionary learning and sparse coding on aggregated N-fMRI datasets from three groups of subjects under each video stimulus, respectively. In our framework, the whole-brain fMRI signals of each subject for one video are extracted and stacked into a 2D matrix S_i , in which a column represented the signal for a voxel. Then signal matrices from all the subjects for one video are pooled and arranged into a big signal matrix S , which consists of three groups of subjects:

$$S = [S_{G_{CD}}, S_{G_{HC}}, S_{G_{PG}}], S_{G_{CD}} = [S_{CD1}, \dots, S_{CDk}], S_{G_{HC}} \\ = [S_{HC1}, \dots, S_{HCk}], S_{G_{PG}} = [S_{PG1}, \dots, S_{PGk}] \quad (4)$$

$$A = [A_{G_{CD}}, A_{G_{HC}}, A_{G_{PG}}], A_{G_{CD}} = [A_{CD1}, \dots, A_{CDk}], A_{G_{HC}} \\ = [A_{HC1}, \dots, A_{HCk}], A_{G_{PG}} = [A_{PG1}, \dots, A_{PGk}] \quad (5)$$

By applying the effective online dictionary learning method on the input fMRI signals matrix, we can learn a dictionary matrix D shared by all three groups of subjects and associated coefficient matrix A that inherit the spatial voxel organization and group correspondence of signal matrix S as Eq. (5), which provides us an effective way to conduct group-wise statistical analysis and comparison. Specifically, each column of D is a representative fMRI signal pattern, and each row in A_x corresponds to the associated coefficient vector that is the spatial distribution among the whole-brain voxels of corresponding

dictionary atom and can be mapped back to brain volume image (A_x represents A_{CDk} , A_{HCk} or A_{PGk} called individual coefficient matrix).

Group-wise statistical coefficient maps

As previous studies applying sparse representation method to single brain fMRI signals, establishing the correspondences of different dictionary atoms across subjects and groups is still an open problem. Although methods such as template matching have been adopted to establish group correspondence, the brain networks analyses for groups are constrained by a limited number of functional brain network templates. In this paper, we employed a group-wise sparse representation method to the N-fMRI datasets of multiple groups of subjects (CD, HC, PG), which can automatically establish group correspondence of different dictionary atoms. Since the dictionary learning and sparse coding strategy does not change the spatial organization and group correspondence of input signal matrix S , the weight coefficient matrix A will maintain the spatial voxel organization as well as group information of S , which provides us a convenient way to conduct group statistical analysis to assess group differences in functional brain networks (Lv et al. 2015c), which is quite different from previous single subject sparse coding. Therefore, the weight coefficient matrix A consists of 3 matrices corresponding to 3 groups, and the matrix corresponding to each group is composed of sub-matrices of subjects (Fig. 1e).

Since the same common mask is adopted to extract fMRI signals of each subject, each voxel has group correspondence among all the subjects, e.g., the reference coefficient of j^{th} voxel to i^{th} dictionary atom in D is preserved in element (i, j) of each sub-matrix (Fig. 1e). For each group respectively, we hypothesize that each coefficient element (i, j) is group-wisely null, and the T-test method is conducted with T defined in Eq. (6) to test acceptance or rejection of null hypothesis for each element (i, j) . The threshold is set to be $P < 0.01$. The derived T-value is transformed to the standard z-score (Friston et al. 1994).

$$T(i, j) = \frac{A_{G_x}(\widehat{l}, j)}{\sqrt{\text{Var}(A_{G_x}(\widehat{l}, j))}}, A_{G_x}(\widehat{l}, j) \\ = \sum_{n=1}^{xk} A_n(i, j), (x = CD \text{ or } HC \text{ or } PG) \quad (6)$$

Here, we select $P < 0.01$ ($Z > 2.3$) as the threshold without multiple comparison correction, which is relatively lower than traditional activation analysis. That's because the online dictionary learning and sparse coding algorithm (Mairal et al. 2010) results in the sparsity of weight coefficient matrix A ,

and if one network is not significantly consistent the coefficient is punished to be zero, which is a strict false positive control. Thus, with a relative low but meaningful threshold, we could possibly detect accurate network spatial maps. Also, the t-test result of A_{GX} is sparse, in which the row represents the non-distribution to the whole brain of each dictionary atom (Fig. 1e). Thus each row can be projected to brain volume image and indicates the distribution of the associated dictionary atom. We define the distribution map of each dictionary atom as a component network. In order to make each component network comparable, we transform the T-value to the standard z-score, color code the z-scores of each component, and then derive the statistical coefficient map. The z-scores can reflect the significance of the contribution of each network. For each type of video stimuli, we conduct the t-tests for A_{GCD} , A_{GHC} , and A_{GPG} , separately. However, the derived statistical coefficient maps for three groups of subjects have group-wise correspondence of the same dictionary atom, thus can be compared cross groups to assess functional brain activities in different brain conditions (Fig. 1e).

Results

In our study, participants viewed three types of videos depicting cocaine, gambling, and sad stories. Cocaine and gambling stories are employed to induce urges for cocaine and gambling in CD and PG groups, respectively. We adopt sad stories video as an emotional active control condition and examine the group differences in emotional processing among three groups of subjects. First, we detect diverse meaningful brain networks in three groups via our group-wise sparse representation method under three types of naturalistic stimuli. In total, 45 functional networks of interest are identified from all the groups under three types of stimuli. Specifically, 15 functional networks are generated from three groups of subjects under sad stories stimuli, and 16 networks are detected under cocaine-use video stimuli. Moreover, 14 functional networks of interest are identified under gambling-use video stimuli. Specifically, the dominant meaningful networks detected in HC, CD and PG groups corresponding to the same dictionary atom have the similar active patterns under each type of video stimuli, respectively. We postulate that CD, PG and HC exhibit similar composition patterns in corresponding networks. Our method characterizes several common brain networks reported previously under each video stimuli, including auditory (sad stories #32, #122; cocaine stories #179; gambling stories #21, #154), visual (sad stories #35, #146; cocaine stories #132, #133; gambling stories #13, #114, #132), and default mode network (sad stories #153; gambling stories #188). Moreover, since the input stimuli of video viewing are quite complex, our method can also identify

some meaningful networks with several associated brain regions, as discussed in each subsection.

Then, we investigate if our group-wise sparse representation method can effectively detect group differences in brain activities during different types of videos viewing. Cross-group comparisons of statistical coefficient maps for three groups of subjects under three types of naturalistic stimuli are conducted. Specifically, we use experimentally determined threshold $Z > 2.3$ to all the detected networks of three groups and compare their voxel numbers (as a metric of spatial distribution) under each stimuli respectively. Furthermore, to test if the networks in one group that exhibit highest voxel numbers also have highest average z scores, we hence compare the average z scores in these networks after thresholding at $Z > 2.3$ among three groups. Specifically, we conduct two-sample t-tests to statistically compare z scores across groups ($P < 0.05$ with FDR correction) under each stimuli, respectively. The result reveals that identified meaningful networks can be used to discriminate different groups of subjects under different video viewing conditions.

Specifically, by using our proposed group-wise sparse representation framework to the three groups of subjects, we can identify networks of interests during video viewing. We use the experimentally determined threshold $Z > 2.3$ on the statistical coefficient maps defined in Section 2.4, and then we can detect voxels that have significant reference to each dictionary atom. Specifically, in our study, there are two steps of identifying meaningful networks. First, using the threshold of $Z > 2.3$, we calculate the voxel number of each statistical coefficient maps in control group, where the maps with high voxel numbers (top 20 %) are selected. Then, we carefully inspect each network component with high voxel number, where components with meaningful brain regions are chosen whereas components without meaningful pattern are discarded. All the manual steps are conducted by voting of at least three researchers (SFig. 1). Thus, we identify several meaningful networks under three types of videos, as shown in the following subsections.

Sad stories

Figure 2a shows the selected 15 most dominant meaningful networks from three groups under sad stories video viewing, which shows similar spatial patterns in corresponding networks across groups. Besides the networks commonly detected under three types of video, our method also identify some meaningful networks composed of several activated regions under sad stories stimuli, such as default mode and executive network (#19), superior temporal gyrus and associated visual cortex (#23), cerebellum, middle and superior temporal gyrus, cingulate and insular cortex (#33), superior temporal gyrus and visual cortex (#47), visual cortex, prefrontal cortex and anterior cingulate cortex (#52), visual and prefrontal cortex

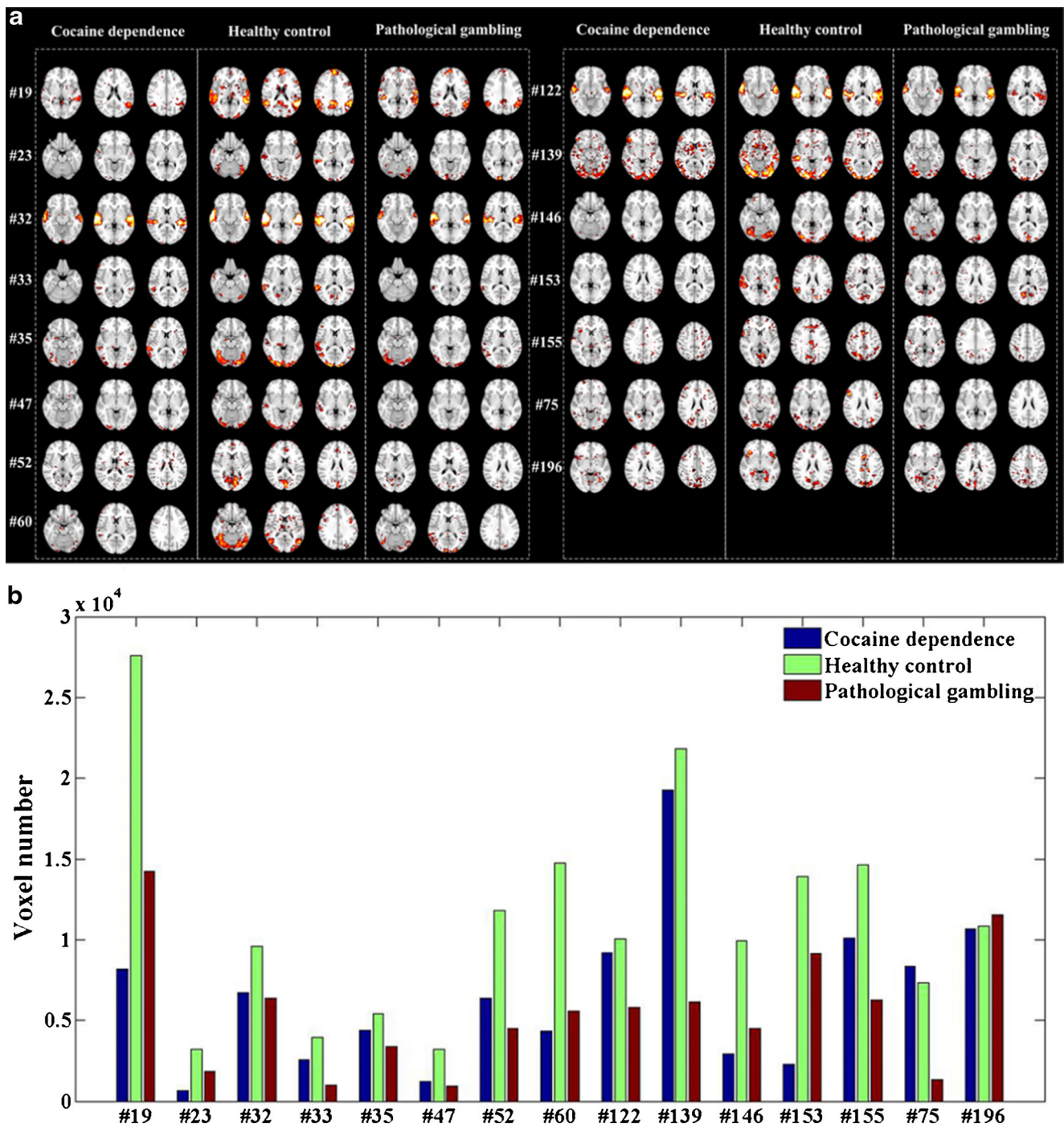


Fig. 2 **a** The statistical coefficient map (z-score map) comparison of 15 networks from three groups under sad stories video stimuli. **b** Voxel number ($P < 0.01$, $Z > 2.3$) comparison of these 15 related networks from three groups

(#60), visual, fusiform and prefrontal cortex (#75), putamen, visual and middle temporal gyrus (#139), visual, motor and anterior cingulate cortex (#155), insular, anterior and posterior cingulate, and somatosensory association cortex (#196).

To investigate the responses of different subjects to sad stories video stimuli, we compare the voxel numbers of statistical coefficient maps across three groups as shown in Fig. 2b. The results indicate that most of networks' region

sizes and functional activations decrease in the PG and CD groups in comparison with HC group. There are only two exceptions, network #75 and #196, where the voxel number of HC decreases compared with CD and PG. Furthermore, we also compare the average z scores in 15 selected networks of three groups, using the same threshold $Z > 2.3$ (SFig. 2(a)), and find that in all the networks average z scores of HC are significantly higher than that of CD and PG.

Cocaine stories

Figure 3a illustrates 16 most dominant meaningful networks from three groups under cocaine stories video viewing. The similar patterns are adopted in corresponding networks across groups. Besides the networks commonly detected under three types of video, there are some additional meaningful networks

identified under cocaine stories stimuli, such as, part of cingulate cortex (#130), visual, superior and middle temporal gyrus, and cerebellum (#14), visual, cerebellum and auditory cortex (#26), visual and fusiform gyrus (#31), visual, auditory and insular cortex (#33), somatosensory cortex and supramarginal gyrus (#64), auditory and middle temporal gyrus (#85), visual, auditory, anterior cingulate and prefrontal cortex (#117),

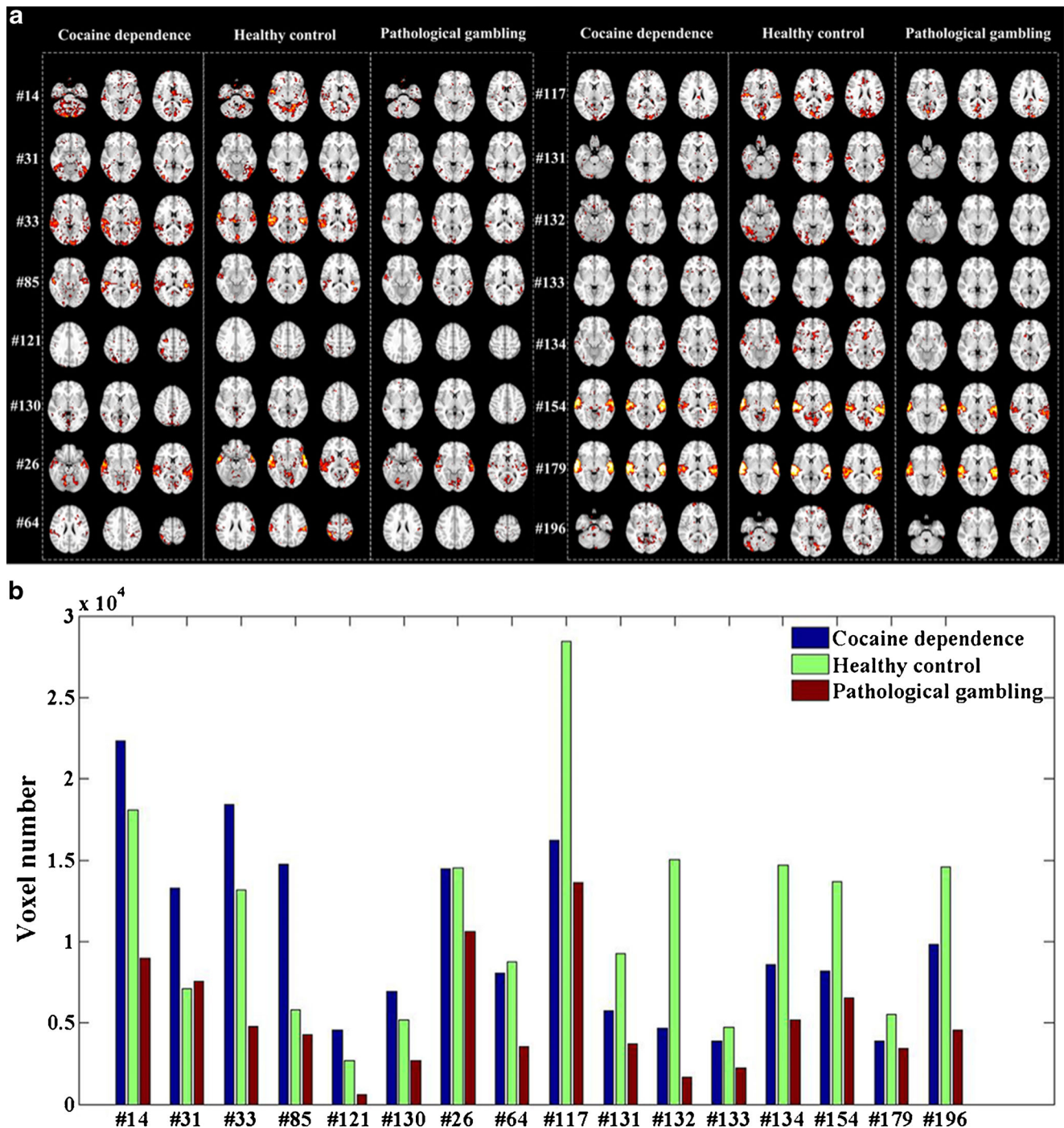


Fig. 3 **a** The statistical coefficient map (z-score map) comparison of 16 networks from three groups under cocaine-use video stimuli. **b** Voxel number ($P < 0.01$, $Z > 2.3$) comparison of these 16 related networks from three groups

premotor, primary motor and somatosensory association cortex (#121), cerebellum, superior and middle temporal gyrus (#131), middle and superior temporal cortex, as well as anterior prefrontal cortex, and part of caudate (#134), auditory and cingulate cortex (#154), cerebellum, visual and prefrontal cortex (#196).

Using the same threshold of $Z > 2.3$, we compare the voxel numbers and active region sizes of all the networks across three groups (Fig. 3(b)). The results can be sorted into two categories. In the first category, CD group exhibits larger numbers of voxels and active regions compared to HC and PG groups, where in most of cases PG group possesses the least number of voxels (network #14, #31, #33, #85, #121, #130). Moreover, we compare the average z scores in these 6 networks thresholded at $Z > 2.3$, and find that average z scores of CD are significantly higher than that of HC and PG, while the HC group has intermediate average z scores and PG group has the lowest (SFig. 2(b)). In the other category, detected brain networks of HC have the largest numbers of voxels and the largest active regions, and the second place is CD, while networks of PG have the least number of voxels and the smallest active brain region (network #26, #64, #117, #131, #132, #133, #134, #154, #179, #196).

Gambling stories

Figure 4a shows the 14 most dominant brain networks detected by our method from three groups during gambling stories video viewing, where similar patterns were detected across groups. Besides the networks commonly detected under three types of video, we can also identify some additional meaningful networks under gambling stories stimuli, such as, superior temporal gyrus (#26), visual, fusiform and premotor cortex (#1), visual, insular and prefrontal cortex (#27), executive control network and posterior cingulate, supramarginal and somatosensory association cortex (#84), executive control network, posterior cingulate and somatosensory association cortex (#99), visual and a part of executive control network (#103), anterior cingulate, somatosensory and supramarginal gyrus (#112), visual, fusiform, insular and auditory cortex (#169).

Using the same experimentally determined threshold $Z > 2.3$ to three groups, we compare the voxel numbers of 14 defined networks across three groups (Fig. 4(b)). There are two categories of results. In the first category, PG group has the largest numbers of voxels and active regions across three groups, where in all cases CD group possesses the lowest number of voxels (network #1, #26, #99). Furthermore, we compare the average z scores in these 3 networks using the threshold of $Z > 2.3$ and find that in these 3 networks average z scores of PG are significantly higher than that of HC and CD, while the HC group has intermediate average z scores and CD group has the lowest (SFig. 2(c)). In the second category, brain

networks of HC possess the largest numbers of voxels and the largest activity brain regions, while networks of CD have the least number of voxels and the smallest active brain region (network #13, #21, #27, #84, #103, #112, #114, #132, #154, #169, #188).

Comparative analysis with tensor independent component analysis

To evaluate and validate the results of our group-wise sparse representation method, we compare it with tensor ICA method implemented in FSL MELODIC toolbox (Beckmann and Smith 2005), which is widely used for decomposing the input data into independent components where stimulus paradigm is consistent among subjects. We conduct tensor ICA for all three groups subjects together under each video stimulus, respectively. The number of components in our study is automatically estimated and optimized by MELODIC toolbox. Here, the estimated numbers of components under sad stories, cocaine stories and gambling stories stimuli are 24, 19 and 40 respectively. Then we adopt dual regression to project tICA components to each subject space, which are then used for calculating group-wise statistical coefficient maps for tensor ICA. Same method is adopted to conduct group-wise statistical analyses.

Supplemental Fig. 3 illustrates the spatial distribution and comparison of voxel number of 8 networks generated from three groups under sad stories stimuli, where all the networks in HC group show highest voxel number. Under cocaine-use video stimuli 7 detected networks and the voxel number comparison results are shown in supplemental Fig. 4, where HC group has highest voxel number, while CD group has intermediate and PG group has the lowest. Moreover, supplemental Fig. 5 shows 7 functional networks of interest identified under gambling-use video stimuli and the comparison of voxel numbers of these networks. PG group shows highest number of voxel in network #7 compared with HC and CD group, and in the remaining identified networks, HC group has highest voxel number, while PG group performs intermediately and CD group has the lowest.

In summary, similar results as our group-wise sparse representation method are achieved, however, the networks detected by our method exhibit more diverse patterns. Under each stimulus, identified networks of three groups have the similar active patterns, but the voxel number of networks can be affected by different brain conditions.

Discussion

Our novel group-wise dictionary learning and sparse representation framework provides group correspondences established

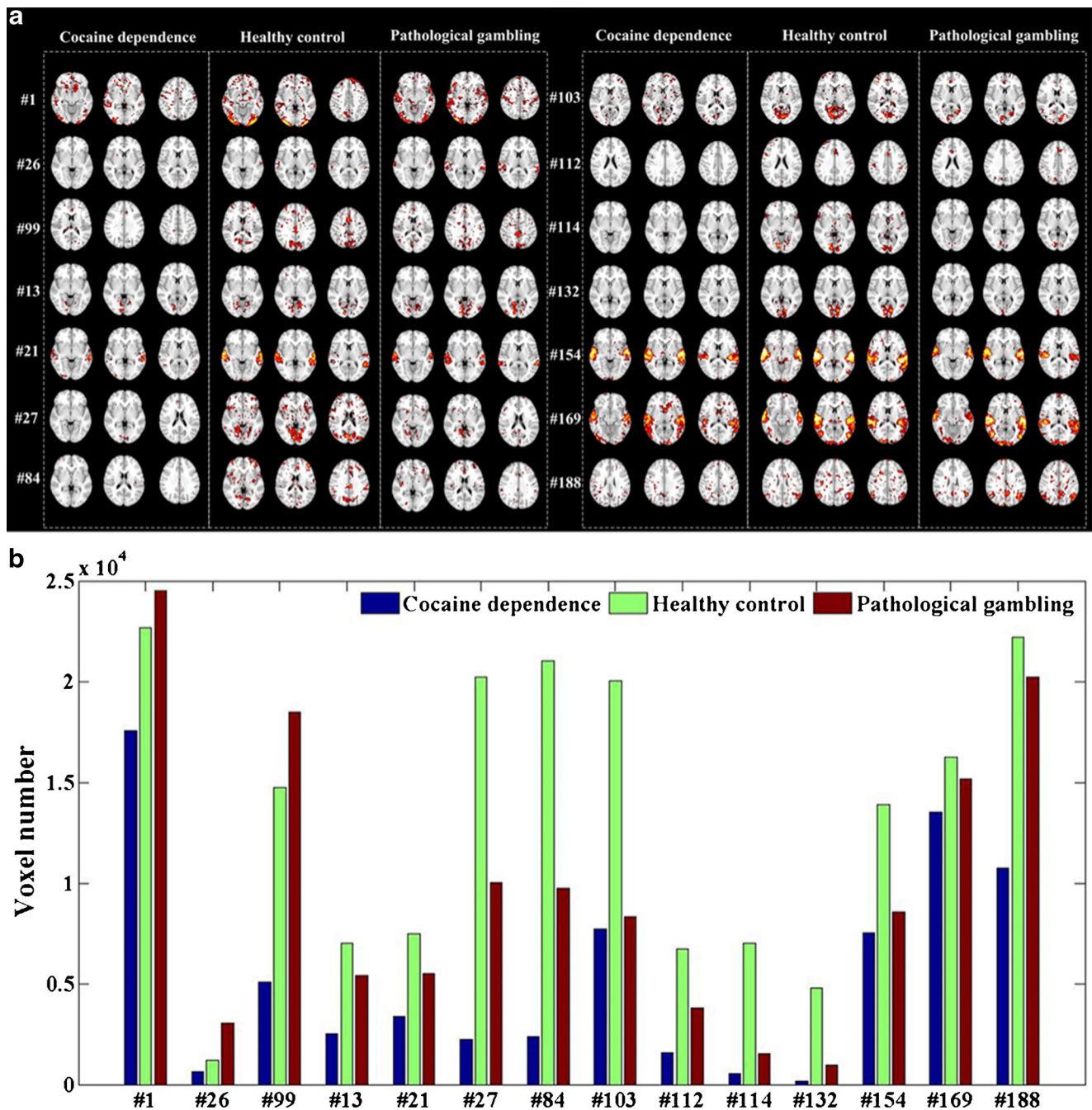


Fig. 4 **a** The statistical coefficient map (z-score map) comparison of 14 networks from three groups under gambling stories video stimuli. **b** Voxel number ($P < 0.01$, $Z > 2.3$) comparison of these 14 related networks from three groups

by the common dictionary to compare the inter-group differences under different types of videos stimuli. Relative to HC, CD and PG patients show weaker response to “natural reward”, that is, sad story. This may be caused by lower dopamine response of CD and PG patients (cite). In addition, CD patients show greater response to cocaine cue in some brain regions relative to PG and HC subjects, while PG patients show greater response to gambling cue in some brain regions compared with CD and HC groups. Our approach can identify brain regions involved in urge of cocaine for CD patients and

urge of gambling for PG patients, including insula, anterior cingulate cortex, prefrontal cortex, and posterior cingulate cortex (Wexler et al. 2001; POTENZA et al. 2003; Crockford et al. 2005; Garavan 2010; Goldstein et al. 2007; Childress et al. 1999; Potenza 2008; Goudriaan et al. 2010). Specifically, insula cortex is directly involved in cocaine craving and gambling urges (Wexler et al. 2001; Garavan 2010; Goudriaan et al. 2010). Similarly, anterior cingulate cortex has been shown to activate during cocaine craving and gambling urges (Wexler et al. 2001; Goldstein et al. 2007; Childress et al. 1999;

Potenza 2008). Also, studies show that decreased activation can be detected in anterior cingulate cortex and prefrontal cortex when cocaine users engaged in GO-NOGO response inhibition task with working memory demands (Hester and Garavan 2004), while both dorsal and ventral regions of mPFC/ACC exhibit relatively more activation during craving in men with CD compared to PG (Potenza 2008). In addition, studies have observed that PG male subjects shows increased activation to video and picture gambling cues in posterior cingulate cortex and dorsolateral PFC (Crockford et al. 2005; Goudriaan et al. 2010). Also, there are studies showing that posterior cingulate activation during viewing of cocaine video is associated with treatment outcome in CD subjects, with those who are able to abstain showing greater activation of this brain region (Kosten et al. 2006).

Moreover, some active brain regions commonly detected by our method under three types of stimuli are involved in emotional stimuli processing. For instance, the anterior cingulate cortex is involved in rational functions, such as empathy and emotion (Decety and Jackson 2004; Jackson et al. 2006). Posterior cingulate cortex (PCC) is associated with emotion and memory, especially in pain and episodic memory retrieval (Nielsen et al. 2005). The superior temporal gyrus is involved in the perception of emotions in facial stimuli (Bigler et al. 2007). Also, insular cortex, which has increasingly drawn attention for its role in subjective emotional experience, plays an important role in mapping visceral states that are associated with emotional experience, giving rise to conscious feelings (Damasio et al. 2000). Notably, since the naturalistic stimuli are more complex and dynamic than the traditional block-based task input, besides several basic functional networks such as visual, auditory and etc., most of the networks detected by our method under each natural stimuli are complex networks with several associated brain regions, which may indicate interactions between primary functional brain regions like visual or auditory cortex and higher functional brain regions like ACC or PCC. The results indicated that our framework can effectively assess group differences in brain activity patterns across different brain conditions (CD, HC, and PG). Although the underlying mechanism is still under exploration, these results provide novel clues for the group differences of brain activities in PG and CD under emotional stimuli.

Sad stories

Based on our results, 15 functional networks are generated from three groups of subjects under sad stories stimuli. In summary, the brain regions involved in emotional stimuli processing, including anterior cingulate cortex, posterior cingulate cortex, superior temporal gyrus, insular cortex and prefrontal regions show decreased voxel numbers and decreased average z-scores in both CD and PG groups (network #19,

#23, #33, #47, #52, #60, #153, #155) (Fig. 2, SFig. 2 (a)). Moreover, HC group shows greater activation in brain regions including amygdala and fusiform gyrus (#35), which are involved in emotion and face processing, and also has greater responses in default mode and executive network (#19), which is associated with working memory performance (Wolf et al. 2015). The prefrontal cortex detected in our results is consistent with results in the previous study (Kober et al. 2016), but they exhibit larger group differences across CD, PG, and HC under emotional stimuli. Interestingly, the fact that some brain regions with decreased voxel numbers and average z scores mentioned above are involved in emotion processing, may indicate that CD and PG subjects tend to have weaker responses/reactions to emotional stimuli compared with HC.

Cocaine stories

In our study, 16 most dominant meaningful networks are identified from three groups under cocaine stories video viewing. Note that it has been reported that brain regions including insula (#33), anterior cingulate cortex (#117), and prefrontal cortex (#117) identified by our method can reflect the urge of cocaine for cocaine dependence individual (Wexler et al. 2001).

Furthermore, the voxel numbers and active region sizes differ across three groups under cocaine stories video viewing as shown in Fig. 3b. The results can be sorted into two categories. In the first category, CD group possesses larger numbers of voxels and active regions compared to HC and PG groups, where in most of cases PG group possesses the least number of voxels (network #14, #31, #33, #85, #121, #130). In addition, in these networks CD group also has the highest average z scores, while HC group performs intermediately, and PG group has the lowest, in all six networks (SFig. 2(b)). The insula cortex (#33) is involved in urge of cocaine for cocaine dependence individual (Wexler et al. 2001; Garavan 2010), other regions such as cingulate and superior temporal gyrus are engaged in emotion perception, formation and processing (Decety and Jackson 2004; Bigler et al. 2007; Damasio et al. 2000), and regions like fusiform, premotor, primary motor and somatosensory association cortex, are responsible for movements planning and execution. In the second category, detected brain networks of HC have the largest numbers of voxels and active regions, and the second place is CD, while networks of PG have the least number of voxels and the smallest active brain region (network #26, #64, #117, #131, #132, #133, #134, #154, #179, #196). Most of these brain regions correspond to basic functional regions, such as auditory and visual cortex, while there are also regions related to visuo-motor coordination, such as somatosensory and supramarginal cortex (#64). Therefore, we can infer that

CD subjects have stronger responses to cocaine-use video stimuli compared with PG subjects, and in some specific emotion and motor related brain regions CD subjects also have stronger activations compared with HC subjects.

Gambling stories

In summary, we identify 14 most dominant brain networks detected by our method from three groups during gambling stories video viewing. Interestingly, the identified brain networks including posterior cingulate cortex (#84, #99) and anterior cingulate cortex (#112) are involved in the urge of gambling for pathological gambling individual (POTENZA et al. 2003; Crockford et al. 2005).

Furthermore, the voxel numbers and active region sizes differ across three groups under gambling stories video viewing as shown in Fig. 4. As we discuss above, there are two categories of results. In the first category, PG group has the largest numbers of voxels and the largest active regions across three groups, where in all cases CD group possesses the least number of voxels (network #1, #26, #99). We also find that PG group has the highest average z scores in these three networks, while HC group has intermediate average z score and CD group has the lowest (SFig. 2(c)). The posterior cingulate cortex (#99) is involved in urge of gambling for pathological gambling individual (POTENZA et al. 2003; Crockford et al. 2005; Goudriaan et al. 2010), other regions such as superior temporal gyrus and executive control network are involved in emotion perception, formation and processing (Decety and Jackson 2004; Bigler et al. 2007), and regions like fusiform, premotor, and somatosensory association cortex, are related to movements planning and execution (Gazzaniga 2004). Therefore, we conclude that in these regions PG group may have the strongest functional activation among three groups. In the second category, brain networks of HC possess the largest numbers of voxels and active brain regions, while networks of CD have the least number of voxels and the smallest active brain regions (network #13, #21, #27, #84, #103, #112, #114, #132, #154, #169, #188). Overall, we can infer that PG subjects have stronger response to gambling video stimuli compared with CD subjects, and in some specific brain regions associated with emotion and motor PG subjects also have stronger activations compared with HC subjects.

In summary, based on comparisons of voxel size and average z score in detected brain networks, we can find that the different brain conditions may lead to different degrees of decrease of active region size and z score in response to different naturalistic stimuli. The effect needs more future interpretation, but it is promising that these interesting findings can be captured by our group-wise sparse representation method.

Conclusion

From a technical perspective, in this paper, we presented a novel group-wise sparse representation and statistical coefficient mapping method to assess group differences in functional brain activations across CD, HC, and PG groups under different naturalistic stimuli. The basic idea of our computational framework is to learn a common time series dictionary matrix from the aggregated N-fMRI signals of all three groups of subjects, and subsequently assess weight coefficient matrices corresponding to each common dictionary atom for each group separately. The methodological advantages of our group-wise sparse representation and statistical coefficient mapping study are summarized as follows. First, compared with widely used model-driven method GLM and data-driven method ICA, our sparse representation framework not only does not need any priori hypothesis of possible causes of brain responses, but also takes the intrinsic sparsity of the whole-brain fMRI signals into consideration. Moreover, the sparse representation of whole-brain fMRI signals can effectively and robustly identify concurrent functional networks for all of task-evoked fMRI, resting-state fMRI (Lv et al. 2015a; Lv et al. 2015b) and natural stimulus fMRI data. Second, compared with previous sparse representation of single subject fMRI signals (Lv et al. 2015a; Lv et al. 2015b), our method can automatically establish the correspondences across individuals and populations of different dictionary atoms, which benefits the group statistical comparison analysis. However, for single subject sparse representation method, the correspondence for the atoms related to task paradigm was established by time-frequency analysis, otherwise the template matching method was adopted to establish correspondence across individuals for resting-state fMRI, which is constrained by the limited number of brain network templates. Third, when learning coefficient, the sparsity constraint regularizes the regressor selection, consequently the results from group non-zero t-test will be much stricter. As a result, statistical coefficient maps are more reliable in measuring the significance of contribution. Finally, compared with task-based fMRI, naturalistic fMRI paradigms use more complex and dynamic stimuli to examine neural processes under more real-life condition. As far as we know, our study is one of earliest applications of applying group-wise sparse representation method to N-fMRI datasets of brain conditions. However, there are also challenges and weaknesses associated with our studies. Since the naturalistic fMRI signals have complex and abundant stimuli, using group-wise sparse representation method, we can learn and optimize hundreds of interesting brain networks, and consequently identify group differences across different brain conditions under different natural stimuli. However, due to the lack of ground truth in fMRI studies, it is difficult to interpret the neural meaning of all the learned brain networks

and group differences across brain conditions. Thus, more frequency, temporal and spatial characterization methods should be developed in the near future for better interpreting our results. Finally, we should also apply this novel framework in other naturalistic fMRI datasets to examine its reproducibility and robustness.

From a clinical perspective, our experimental results demonstrated that group-wise sparse representation method can detect multiple meaningful brain networks concurrently, and these networks consistently exist across three groups under three types of videos, despite that obvious group differences in functional activations can be revealed with different brain conditions of CD, HC, and PG. Although some results and underlying mechanisms are beyond our current explanation, they essentially provide novel, important clues for the effects of CD and PG on the human brain.

Compliance with ethical standards

Funding This study was funded by NSF CAREER Award (IIS-1,149,260), NIH R01 DA-033,393, NIH R01 AG-042,599, NSF CBET-1,302,089, NSF BCS-1,439,051 and Fundamental Research Funds for the Central Universities 3102015ZY046.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent: Informed consent was obtained from all individual participants included in the study.

References

- Baron-Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., et al. (1999). Social intelligence in the normal and autistic brain: an fMRI study. *European Journal of Neuroscience*, *11*(6), 1891–1898.
- Bartels, A., & Zeki, S. (2004). The chronoarchitecture of the human brain—natural viewing conditions reveal a time-based anatomy of the brain. *NeuroImage*, *22*(1), 419–433.
- Beckmann, C. F., & Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject fMRI analysis. *NeuroImage*, *25*(1), 294–311.
- Bigler, E. D., Mortensen, S., Neeley, E. S., Ozonoff, S., Krasny, L., Johnson, M., et al. (2007). Superior temporal gyrus, language function, and autism. *Developmental Neuropsychology*, *31*(2), 217–238.
- Bordier, C., Puja, F., & Macaluso, E. (2013). Sensory processing during viewing of cinematographic material: Computational modeling and functional neuroimaging. *NeuroImage*, *67*, 213–226.
- Childress, A. R., Mozley, P. D., McElgin, W., Fitzgerald, J., Reivich, M., & O'Brien, C. P. (1999). Limbic activation during cue-induced cocaine craving. *American Journal of Psychiatry*, *156*, 11–18.
- Codispoti, M., Surcinelli, P., & Baldaro, B. (2008). Watching emotional movies: Affective reactions and gender differences. *International Journal of Psychophysiology*, *69*(2), 90–95.
- Crockford, D. N., Goodyear, B., Edwards, J., Quickfall, J., & el-Guebaly, N. (2005). Cue-induced brain activity in pathological gamblers. *Biological Psychiatry*, *58*(10), 787–795.
- Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J., et al. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience*, *3*(10), 1049–1056.
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, *2*(4), 189–210.
- Garavan, H. (2010). Insula and drug cravings. *Brain Structure and Function*, *214*(5–6), 593–601.
- Gazzaniga, M. S. (2004). *The cognitive neurosciences*. Cambridge, MA: MIT press.
- Goldstein, R. Z., Tomasi, D., Rajaram, S., Cottone, L. A., Zhang, L., Maloney, T. E., et al. (2007). Role of the anterior cingulate and medial orbitofrontal cortex in processing drug cues in cocaine addiction. *Neuroscience*, *144*(4), 1153–1159.
- Goudriaan, A. E., De Ruiter, M. B., Van Den Brink, W., Oosterlaan, J., & Veltman, D. J. (2010). Brain activation patterns associated with cue reactivity and craving in abstinent problem gamblers, heavy smokers and healthy controls: an fMRI study. *Addiction Biology*, *15*(4), 491–503.
- Gur, R. C., Alsop, D., Glahn, D., Petty, R., Swanson, C. L., Maldjian, J. A., et al. (2000). An fMRI study of sex differences in regional activation to a verbal and a spatial task. *Brain and Language*, *74*(2), 157–170.
- Han, J., Ji, X., Hu, X., Guo, L., & Liu, T. (2015). Arousal recognition using audio-visual features and fmri-based brain response. *IEEE Transactions on Affective Computing*, *6*(4), 337–347.
- Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48.
- Heatherington, T. F., Kozłowski, L. T., Frecker, R. C., & FAGERSTROM, K. O. (1991). The Fagerström test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. *British Journal of Addiction*, *86*(9), 1119–1127.
- Hester, R., & Garavan, H. (2004). Executive dysfunction in cocaine addiction: evidence for discordant frontal, cingulate, and cerebellar activity. *The Journal of Neuroscience*, *24*(49), 11017–11022.
- Hu, X., Lv, C., Cheng, G., Lv, J., Guo, L., Han, J., et al. (2015). Sparsity-constrained fMRI Decoding of visual saliency in naturalistic video streams. *IEEE Transactions on Autonomous Mental Development*, *7*(2), 65–75.
- Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia*, *44*(5), 752–761.
- Kober, H., Lacadie, C. M., Wexler, B. E., Malison, R. T., Sinha, R., & Potenza, M. N. (2016). Brain activity during cocaine craving and gambling urges: an fMRI study. *Neuropsychopharmacology*, *4*(2), 628–637.
- Kosten, T. R., Scanley, B. E., Tucker, K. A., Oliveto, A., Prince, C., Sinha, R., et al. (2006). Cue-induced brain activity changes and relapse in cocaine-dependent patients. *Neuropsychopharmacology*, *31*(3), 644–650.
- Kret, M. E., & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia*, *50*(7), 1211–1221.

- Lahnakoski, J. M., Glerean, E., Salmi, J., Jääskeläinen, I. P., Sams, M., Hari, R., et al. (2012). Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Frontiers in Human Neuroscience*, *6*, 233.
- Lesieur, H. R., & Blume, S. B. (1987). The South Oaks Gambling Screen (SOGS): a new instrument for the identification of pathological gamblers. *American Journal of Psychiatry*, *144*(9), 1184–1188.
- Lv, J., Jiang, X., Li, X., Zhu, D., Chen, H., Zhang, T., et al. (2015a). Sparse representation of whole-brain fMRI signals for identification of functional networks. *Medical Image Analysis*, *20*(1), 112–134.
- Lv, J., Jiang, X., Li, X., Zhu, D., Zhang, S., Zhao, S., et al. (2015b). Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function. *IEEE Transactions on Biomedical Engineering*, *62*(4), 1120–1131.
- Lv, J., Jiang, X., Li, X., Zhu, D., Zhao, S., Zhang, T., et al. (2015c). Assessing effects of prenatal alcohol exposure using group-wise sparse representation of fMRI data. *Psychiatry Research: Neuroimaging*, *233*(2), 254–268.
- Mairal, J., Bach, F., Ponce, J., & Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *The Journal of Machine Learning Research*, *11*, 19–60.
- Manoach, D. S., Gollub, R. L., Benson, E. S., Searl, M. M., Goff, D. C., Halpern, E., et al. (2000). Schizophrenic subjects show aberrant fMRI activation of dorsolateral prefrontal cortex and basal ganglia during working memory performance. *Biological Psychiatry*, *48*(2), 99–109.
- Nielsen, F. Å., Balslev, D., & Hansen, L. K. (2005). Mining the posterior cingulate: segregation between memory and pain components. *NeuroImage*, *27*(3), 520–532.
- Pessoa, L. (2012). Beyond brain regions: network perspective of cognition–emotion interactions. *Behavioral and Brain Sciences*, *35*(03), 158–159.
- Potenza, M. N. (2008). The neurobiology of pathological gambling and drug addiction: an overview and new findings. *Philosophical Transactions of the Royal Society of London B. Biological Sciences*, *363*(1507), 3181–3189.
- Potenza, M. N., Steinberg, M. A., Skudlarski, P., Fulbright, R. K., Lacadie, C. M., Wilber, M. K., et al. (2003). Gambling urges in pathological Gambling: A functional magnetic resonance imaging study. *Archives of General Psychiatry*, *60*(8), 828–836.
- Tapert, S. F., Brown, G. G., Kindermann, S. S., Cheung, E. H., Frank, L. R., & Brown, S. A. (2001). fMRI measurement of brain dysfunction in alcohol-dependent young women. *Alcoholism: Clinical and Experimental Research*, *25*(2), 236–245.
- Wexler, B. E., Gottschalk, C. H., Fulbright, R. K., Prohovnik, I., Lacadie, C. M., Rounsaville, B. J., et al. (2001). Functional magnetic resonance imaging of cocaine craving. *American Journal of Psychiatry*, *158*(1), 86–95.
- Wolf, D. H., Satterthwaite, T. D., Calkins, M. E., Ruparel, K., Elliott, M. A., Hopson, R. D., et al. (2015). Functional neuroimaging abnormalities in youth with psychosis spectrum symptoms. *JAMA Psychiatry*, *72*(5), 456–465.