HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2019 April 19.

Published in final edited form as:

Neuroimage. 2017 August 15; 157: 716–732. doi:10.1016/j.neuroimage.2017.06.032.

MAPBOT: Meta-analytic parcellation based on text, and its application to the human thalamus

Rui Yuan[a,b], **Paul A. Taylor**[c], **Tara L. Alvarez**[a], **Durga Misra**[b], and **Bharat B. Biswal**[a,d,*]

[a]Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ 07102, US

[b]Department of Electrical Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA

[c]Scientific and Statistical Computing Core, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, USA

[d]Department of Radiology, Rutgers, The State University of New Jersey, Newark, NJ 07102, USA

Abstract

Meta-analysis of neuroimaging results has proven to be a popular and valuable method to study human brain functions. A number of studies have used meta-analysis to parcellate distinct brain regions. A popular way to perform meta-analysis is typically based on the reported activation coordinates from a number of published papers. However, in addition to the coordinates associated with the different brain regions, the text itself contains considerably amount of additional information. This textual information has been largely ignored in meta-analyses where it may be useful for simultaneously parcellating brain regions and studying their characteristics. By leveraging recent advances in document clustering techniques, we introduce an approach to parcellate the brain into meaningful regions primarily based on the text features present in a document from a large number of studies. This new method is called MAPBOT (Meta-Analytic Parcellation Based On Text). Here, we first describe how the method works and then the application case of understanding the sub-divisions of the thalamus. The thalamus was chosen because of the substantial body of research that has been reported studying this functional and structural structure for both healthy and clinical populations. However, MAPBOT is a general-purpose method that is applicable to parcellating any region(s) of the brain. The present study demonstrates the powerful utility of using text information from neuroimaging studies to parcellate brain regions.

Keywords

Meta-analysis; Thalamus; Text Mining; Topic Mapping

*Correspondence to: New Jersey Institute of Technology, Department of Biomedical Engineering, University Heights, Newark, NJ 07102, USA. bbiswal@gmail.com (B.B. Biswal).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neuroimage.2017.06.032.

Introduction

In the last two decades, neuroimaging researchers have produced an exponentially increasing number of studies localizing activation in specific brain regions in both healthy and diseased populations. Several imaging modalities, such as structural MRI, functional MRI (fMRI) and positron emission tomography, have revealed various perspectives of brain structures and functions while implementing a variety of different experimental designs, group sizes, inclusion criteria, etc. One of the major challenges for neuroimaging researchers has been to synthesize the results from these diverse publications. An initial approach was made by manually grouping studies with a similar topic together and then summarizing the reported activation locations into a table or figure (Buckner and Petersen, 1996; Poeppel et al., 1996; Owen et al., 1997). However, these criteria and techniques were not subject to statistical validation.

In recent years, due to the wide use of standard spatial normalization in group studies, activation locations across different subjects or studies could be reported on the same common template with the corresponding coordinates of interest. Simultaneously, researchers have tried to develop probabilistic approaches for quantifying the uncertainty of the different spatial locations obtained from various studies. This has allowed results to be integrated across studies in a quantitative way using algorithms such as activation likelihood estimation (ALE; Turkeltaub et al., 2012) and multilevel kernel density analysis (MKDA; Wager et al., 2009). Additionally, the Brainmap (Fox and Lancaster, 2002; Laird, 2005) and Neurosynth (Yarkoni,2011) projects provide convenient tools to automatically perform this coordinate based meta-analysis across the neuroimaging literature. Based on a well-studied background of existing standard coordinate systems and mature cluster techniques, meta-analytic connectivity modeling based parcellation (MACM-CBP) has been developed to group voxels into clusters based on the similarity between each voxel’s co-activation maps (Barron et al., 2015Robinson et al., 2015; Chang et al., 2013). This technique has been successfully applied to subdividing brain regions, such as the insula (Chang et al., 2013 Cauda et al., 2012), pulvinar (Barron et al., 2015), temporo-parietal junction (Bzdok et al., 2013) and orbitofrontal cortex (Kahnt et al., 2012).

However, presently available meta-analysis-based parcellation algorithms, such as those listed above, are limited to only using the coordinate data in tables reporting statistically significant locations and clusters. These approaches ignore the remaining text content of the papers, which one could argue comprises the majority of the information of the published work. Indeed, the reader of a paper typically acquires several types of information from the written text about brain regions such as but not limited to the following information: functionality, roles, interregional connectivity, and behavioral associations. This rich information is in addition to the simple reported spatial coordinates found within the tables of papers. It is true that extracting contextualized information among publication text is known to be challenging, recent advances in text mining and nature language processing have provided some effective ways to address these problems. For instance, by utilizing the co-occurrence of individual text terms to produce maps of semantic structures and to provide insights into how knowledge is organized within the large corpus of literature (Beam et al., 2014; Carley et al., 1997; Diesner and Carley, 2005). Moreover, the bag-of-word model can

be used to represent the summation of a document as a group of key words regardless grammar or word order. Therefore, a document can be represented as a frequency vector of feature words. Based on this simplified model, document partition, agglomerative (hierarchical) clustering, and topic mapping can be efficiently performed on massive collection of text data. This conceptual model has been proven to be useful at many situations.

Thus, combining both the existing concepts in standard metaanalysis and the recently available methods from text mining, here we introduce a new meta-analysis-based method to parcellate brain into meaningful regions, called MAPBOT (Meta-Analytic Parcellation Based On Text).. As opposed to several existing methods described above, which are primarily built on the spatial similarity of co-activation maps, our proposed approach is instead driven by contextual similarity (i.e., the relation of co-occurrence of terms) across papers.

In this study, we applied our technique to the literature describing the human thalamus. The thalamus was chosen because it has a widely distributed set of connections among cortical and subcortical regions and appears to be involved with most cognitive functions (Sherman et al., 2006; Sherman and Guillery et al., 2013; Jones, 1998, 2001, 2009). Neuroimaging studies have made significant progress toward advancing our understanding of the human thalamus in vivo by using the diffusion tensor imaging (DTI) (Behrens et al., 2003; Draganski et al., 2008; Traynor et al., 2010; O'Muircheartaigh et al., 2011) and fMRI (Zhang et al., 2008, 2010; Kim et al., 2013; Yuan et al., 2016). These previous studies have investigated the topography of thalamocortical system among distinct thalamic sub-regions with large cortical regions (Zhang et al., 2008, 2010; Behrens et al., 2003) or with networks (Yuan et al., 2016). The dysfunction of the thalamus has been associated with several psychotic disorders, including major depression (Greicius et al., 2007), Parkinson's disease (Fasano et al., 2012), and schizophrenia (Andreasen et al., 1994; Corradi Dell'Acqua et al. 2012; Popken et al., 2000), and a cross modality parcellation map of the thalamus may be useful in understanding thalamic functions and the underlying potential mechanism(s) of these associated diseases.

Despite many research advances to understand thalamic function, several questions about the role of the thalamus in a broad sense of cognitive function remain unclear. Most previous studies have focused separately on either functional connectivity (in particular, using resting state fMRI paradigms) or structural connections. However, no single imaging study can conduct all the possible tasks needed to completely explore all the thalamic functions. Moreover, larger nuclei such as the medial dorsal nucleus and the pulvinar are known to have involved within multiple functions (Barron et al., 2015; Shipp et al., 2003; Yuan et al., 2016). Therefore, the homogeneity of functional distinctions of thalamic subdivisions remains unclear.

This research is organized into the following sections. First, we introduce the text-based parcellation method MAPBOT. The second section applies our method to the existing literature that studies the human thalamus. This allows us to describe the groups of topics that are related specifically to each thalamic sub-division. Last, we summarize results that

demonstrate the usefulness of using MAPBOT within meta-analyses of text features to study brain regions. The future direction of our research is to further enhance the understanding of functional topography of the human thalamus and potentially other brain regions as described in the published literature.

Methods

Materials

To implement the MAPBOT method (Fig. 1), we made use of Neurosynth (Yarkoni et al., 2011), which is an open access dataset which contains a large number of studies that provides raw metadata as well as particular extracted text features. The Neurosynth database was downloaded from (https://github.com/neurosynth/neurosynth-data; the latest update was on July 2015). This database contains over 10,000 papers. Nearly 3109 feature terms have been automatically extracted from their abstracts, including both single words and two grams.

There are two types of data sets contained within the Neurosynth database (Fig. 1). The first data set can be described as the "document by coordinates" (DC) matrix, whose $C_{\mathrm{DC}} = 8$ columns are: pmid, title, year, doi, authors, study index, (x, y, z) spatial coordinates of voxels and the template types (MNI, Talairach, "unknown,"). The number of rows R_{DC} in the matrix is the sum of voxel locations recorded per document (l_i) throughout full corpus of N_{d} documents:

$$R_{DC} = \sum_{i=1}^{N_d} l_i,$$

which is typically a large number (in current dataset, R_{DC}= 386,455).

Traditional MACM is based on this form of data.

The second type of data set in Neurosynth is the "document by terms" (DT) matrix. The number of rows in the matrix is determined by the size of the corpus, such that $R_{\mathrm{DT}} = N_{\mathrm{d}}$ (= 11,406). Each row represents a single study. Each column represents the weights for a different text feature. There are total 3109 text terms, which are automatically extracted the abstract of all documents. Names of all the text terms are in the first row. The term weights are normalized using the term-frequency and inverse document frequency (tf-idf) values. For a given term t_j in document d_i, the term-frequency tf_{ij} is defined as the number of occurrences of t_j in document d_i. The inverse document frequency is defined as

$$idf_j = ln\left(\frac{N_d}{n_j}\right) idf_j = ln\left(\frac{N_d}{n_j}\right),$$

where n_j is the number of documents in which term t_j appears. We note that here, the "entire corpus" refers to the full set of "all documents that have been selected"; which means this could be a specific subset from the full NeuroSynth database according to some systematic

criterion, such as a year range, a certain brain region, a specific journal. Finally, the ij-th matrix element of the DT matrix M is constructed as follows:

$$M_{ij} = tf_{ij} * idf_j,$$

such that each element is effectively the count of terms per document, weighted (logarithmically) by the number of documents in the corpus containing that term.

In this study, we propose a method for combining the information contained in both the DC and DT matrices, with particular interest in the large amount of information represented by the latter. However, the DT matrix cannot be directly used for parcellation in the form given above, but it must instead be transformed, as described below, to construct the voxel by term (VT) matrix. First, based on a user-defined region of interest (ROI) with N_v voxels, all documents that reported a voxel of that region from the DC matrix will be listed. From this set of documents, corresponding rows in the DT matrix are identified and their weights extracted. Thus, for each voxel there are several reported documents, each having a vector of term-weights. The VT matrix is constructed by averaging weights across the documents for each voxel, so that the matrix dimensions are $R_{VT} = N_v$ rows by $C_{VT} = C_{DT} = N_T$ columns.

Clustering method: nonnegative matrix factorization

In practice, a given VT matrix X is typically sparse, since most documents only contain a small subset of archived terms. Therefore, traditional clustering methods such as k-means, hierarchical and spectral clustering may fail to provide robust and accurate performance (Xu et al., 2003; Gills et al., 2014). A popular method widely used in text mining, called the nonnegative matrix factorization (NMF), which was introduced by Lee and Seung (1999), has been shown to overcome problems with sparsity. It has been applied to the analysis of highdimensional data such as high resolution images, text mining, and gene expression (Lee and Seung et al., 1999; Xu et al., 2003; Shahnaz et al., 2006; Pauca et al., 2004; Kim et al., 2003; Brunet et al., 2004). In contrast to other common dimension reduction methods (e.g, independent or principal components analysis), the non-negative constraints of the method enables NMF to allow only additive, not subtractive, combinations of parts-based representations (Lee and Seung et al., 1999). Moreover, NMF can simultaneously parcellate the VT matrix X along two dimensions. This is a benefit for MRI applications because clustering can be simultaneously based on both voxels and text terms. The clustering on voxels can provide a parcellation of a region, and the later one on text terms basically performed a topic modeling. As suggested in previous studies (Xu et al., 2003; Gillis et al., 2014), using a normalized weight vector on the data has a positive effect on the NMF performance, such as improving cluster accuracy based on comparing the clustering results with the manually classified data. Thus, a weighting matrix $D = \mathrm{diag}\left(X^{\mathrm{T}} X e\right), e = [1, 1, \ldots, 1]^{\mathrm{T}}$, where the length of e and D is C_{VT}, is calculated, resulting in the more practically useful $X' = XD^{-1/2}$.

Next, the weight-normalized matrix X' is entered into NMF, which we describe briefly here using a given factorization rank k, the number of clusters to parcel (i.e., the dimensionality

reduction of the N_v voxels). Given such a matrix X' and k, NMF finds non-negative matrices U and V^T, such that

$$X' = UV^T,$$

where U is a $N_v \; x \; k$ matrix whose columns will define parcellation weights for each voxel; and V is a k x N_T matrix whose rows will define term weights for each parcel.

A common way to solve for U and V^T (Lee and Seung et al., 1999) is to minimize the following (scalar) objective function:

$$O = \left|X' - UV^T\right|^2.$$

The objective matrix function O can be re-written as:

$$O = tr\left(X'X'^T\right) - 2tr\left(X'VU^T\right) + tr\left(UV^TVU^T\right),$$

where tr() is the trace of the matrix. Since this a typical constrained optimization problem, we used the Lagrange multiplier α and β method to solve the equation. The Lagrangian was defined as:

$$L = O + tr\left(\alpha U^T\right) + tr\left(\beta V^T\right),$$

and we then calculate the derivatives of L with respect to U_{ij} and V_{ij}. By

using the Kuhn-Tucker condition (derivatives $\frac{\partial L}{\partial U} \wedge \frac{\partial L}{\partial V}$ are equal to zeros; $\alpha U_{ij} = 0$ and $\beta V_{ij}^T = 0$), the updating formulas are:

$$\overleftarrow{U}_{ij} \leftarrow U_{ij}\frac{(XV)_{ij}}{UV^TV_{ij}},$$

and

$$\overleftarrow{V}_{ij} \leftarrow U_{ij}\frac{(XU)_{ij}}{VU^TU_{ij}}$$

Several ways to estimate the optimal solution have been proposed, such as multiplicative updates (MU; Daube-Witherspoon and Muehllehner et al., 1986; Lee and Seung et al., 2001) and alternating nonnegative least squares (Berry et al., 2007). The MU is perhaps the most popular one and was used in this study, because it is easy to implement and well scaled. However, it should be noted that it does tend to converge relative slowly. A possible

alternative strategy would be to update the weight matrix U several times ahead of updating V (Xu et al., 2003; Gills et al., 2014).

Since the NMF is NP-hard in general (Vavasis et al., 2009), for most methods of solving it, the initialization has considerable effect on the results. We utilized nonnegative double single vector decomposition (nndSVD) on the X' matrix to estimate the initial approximation of U (Boutsidis et al., 2008). The nndSVD is better fit for a sparse matrix. If the X' is rather dense, then random numbers (nonnegative, in the interval 0 to mean of X' divided by 100) were used to replace zeros in the nndSVD (Boutsidis et al., 2008; Gillis et al., 2014). Although a simple way to initialize U or V matrix is using random initial values, SVD-based initialization is preferable due to its faster convergence and typically lower error rate once converged (Boutsidis et al., 2008).

When the process reached a specified lambda-tolerance (or a predefined maximum number of iterations), the output solution was obtained by normalizing the latest U and V. Based on the U matrix, each voxel was assigned to a cluster (labeled with a number from 1 to k), which has the largest weight in the corresponding row of U.

As a final step, although this method does not require merging segments or labels, the common technique to eliminate "isolated" voxels was also applied. The cluster label of an isolated voxel was switched to that of the majority of its neighbors (out of 27 nearest neighbors).

Dimensionality reduction specification

Similar to other decomposition schemes such as independent component analysis, the number of components (here, parcellated sub-regions) k must be chosen. The following scheme for determining the final number of regions in a data-driven manner was implemented, utilizing two criteria.

First, the (normalized) variance of information (VI; Meila et al., 2003) was used to measure the similarity between the neighboring cluster counts (i.e., between final numbers of k and $k-1$ clusters). This is measured as:

$$VI\left(C_k, C_{k-1}\right)_k = H\left(C_k\right) + H\left(C_{k-1}\right) - 2MI\left(C_k, C_{k-1}\right),$$

where H represents the entropy of the cluster solution C, and MI represents the mutual information shared by the two cluster solutions. A "good" solution k^* shows a decrease in VI from $k-1$ to k or an increase from k to $k+1$ (called a gap pattern). Such "good" factorization ranks were selected candidates for the final decomposition.

The second criterion for evaluating decompositions was the Hartigan index (HI; Hartigan, 1985). The optimal solution would minimize $|HI(k) - HI(k-1)|$. We consider any cluster K whose $|HI(k) - HI(k-1)|$ shows decrease from k-1 and increase from k+1 could be the potential optimal solution. The HI index is defined as:

$$HI_k = \left(\frac{tr(W_k)}{tr(W_{k+1})} - 1 \right)(n - k - 1)$$

where W_k is the within cluster distance of the cluster k; tr() is the trace of the matrix; and n equals the total number of voxels in the region.

To evaluate the symmetry of the parcellation results, the symmetric index was also estimated by calculating VI values between the left (CL) and right (CR) clustering:

$$symm\_VI(CL_k, CR_k)_k = VI(CL_k, CR_k)_k = H(CL_k) + H(CR_k) - 2MI(CL_k, CR_k),$$

where a smaller value indicates more symmetric clustering results. All metrics were calculated separately for the left and right with different factorization rank k. The "good" k candidates from the left and right were determined respectively by voting from its VI and HI. Then, the candidate with the lowest symmetric VI is to determine the best k^* solution (Fig. 2).

In practice, the user must select minimal and maximal bounds for k within which to search, such that the interval $k_{min} <= k <= k_{max}$ is investigated. These should be chosen within a reasonable, wide interval; if, after iterating through this range, $k^* = k_{min}$ or k_{max}, then likely a wider interval should be selected.

All programs to do the above, comprising MAPBOT, were written in MATLAB (MathWorks Inc.), making use of DC and DT matrices obtained from Neurosynth. AFNI (Cox et al., 1996) functions were used for visualizing the volumetric results as slices, as well as for exporting data via its AFNI_MATLAB package. SUMA (Saad and Reynolds, 2012) was used to visualize the parcellation results as surfaces in 3D space.

Applying MAPBOT: Thalamus parcellation and evaluation metrics

The full thalamus volume was defined using the Morel template (Krauth et al., 2010; Morel et al., 1997), which was down-sampled to 2 mm isotropic voxels to match the MNI space used here. Then, this mask was split along the midline into the left (L-) and right (R-) regions. In order to reduce bias from selecting a specific voxel radius for association, five sets of documents were created based on the thalamus mask, selecting those documents that reported coordinates within 4, 6, 8, 10, and then 12 mm of any thalamic voxel were retrieved from Neurosynth. For each radius-based collection, a DT (tf-idf) matrix was first constructed with idf based on its cache of documents, and then the VT matrix was obtained through the transform mentioned above. The final VT matrix, used in subsequent analysis, is estimated by averaging the five initial VT matrices. The range of possible number of clusters in the thalamus (each hemisphere) was set to $2 <= k <= 20$, so that there were 19 different parcellation results in total from which an optimal solution from the combined VI and HI criteria was obtained.

Sub-division corresponded terms and dual-view topic mapping

As mentioned above, an interesting property of the NMF is a dual view of decomposition $X \sim UV^{T}$. The NMF does not only perform clustering on the rows of VT matrix, defining spatial parcels, but also on its columns, which groups the terms into different sets, commonly known as topic mapping (Fig. 1).

We obtained the V matrix from NMF with each assigned factorization rank k. Each ith thalamic sub-division (described by the ith column of U) has its own set of corresponding terms (described by the *ith row of* V^{T}). Similar to the way that a voxel is assigned to a cluster, each term was assigned to a cluster that had the largest weight at the corresponding column of V^{T}. To further explore the relation with existing topics of neuroimaging interest, a hypergeometric analysis was used to address the possibility of observing the overlaps between the set of terms from each thalamic sub-division, and the set of terms from existing topics described by Poldrack et al. (2012). Based on the number of topics inside each set, we named the sets the topics-50, topics-100, topics-200, and topics-400. For each topic set, the top 5 topics from each thalamic cluster on either hemisphere were presented.

Post-hoc topic mapping of thalamic subdivisions

To better illustrate the specific topics of each thalamic subdivision, the forward inference approach was used to quantitatively characterize the neurofunctional profiles with respect to topics and behavior domains. First, the binomial test of whether P(activation | term) > P(activation) was performed on each term of each topic to test whether the term was "over-represented" by given activation within the cluster (Robinson et al., 2015; Riedel et al., 2015). Second, the modified Fisher's method (Brown et al., 1975; Kost et al., 2002) was used to estimate the probability that each topic was over-represented, by combining the previous term-based tests.

The original Fisher's method assumes the combined term-based tests are independent of each other, but that is not necessarily required to be the case for terms here. Therefore, the modified Fisher's method, which considers the covariance between terms, was used (Brown et al., 1975; Kost et al., 2002). Therefore, significance was assessed using a chi-square measure of corrected combined tests and corrected degree of freedom, and a result was termed "significant" when it passed a Bonferroni corrected $P < 0.05$ threshold.

In addition, a reverse inference approach was conducted in a Bayesian manner. This analysis aimed to determine whether the terms or topics were over-represented for each cluster compared to the terms that were observed in the Neurosynth database. The likelihood of P(term | activation) can be derived from P(activation | term) as well as P(term) and P(activation) through Bayes' rule:

$$P(term \mid activation) = \frac{P(activation \mid term) * P(\text{term})}{P(activation)}.$$

Then, the Bayesian factor was used to address the substantial overrepresentation by estimating the ratio between posterior probability P(Term | Activation) and *a priori* probability P(Activation| term). The top 5 topics from each thalamic cluster were listed.

Resting state functional connectivity

Finally, in order to investigate the thalamocortical connectivity associated with the present parcellation scheme, we utilized the Cambridge data set from the publicly available 1000 Functional Connectomes Project (FCP) database (Biswal et al., 2010). The resting state FMRI data sets were processed according to (Yuan et al., 2016), which we briefly summarize here. Resting-state fMRI and anatomical images were downloaded from the FCP website ((http://fcon_1000.projects.nitrc.org/), consisting of 198 subjects (75 males, 123 females), with ages between 18 and 30 years. The fMRI images were collected using a 3T scanner, with TR of 3 s, 47 slices, 3 mm isotropic voxel size and 119 time points. The first 5 time points were removed, leaving 114 time points for each subject. Each subject's T1-weighted anatomical scan had been acquired using a magnetization-prepared rapid-acquisition gradient echo (MPRAGE) sequence (192 slices with a 144*192 matrix; voxel size = $1.20\times1.00\times1.33$ mm^3). The fMRI datasets were realigned to the first image to correct for head motion and linearly co-registered to each subject's T1-weighted image. The 6 motion parameters obtained from the rigid body registration and the Euclidean norm of all motion derivatives were regressed out from time series of all the voxels. Each structural image was segmented into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF). The functional images were then transformed to the standard Montreal Neurological Institute (MNI) template in $3\times3\times3$ mm^3 by using the Diffeomorphic Anatomical Registration Through Exponentiated Lie algebra (DARTEL, Ashburner, 2007) toolbox. The CSF and WM masks were defined by thresholding individual tissue probability maps at 0.95. The first 5 principal components from each of the CSF and WM masks were regressed out from time series of every voxel. Finally, a bandpass filter ranging between 0.01 and 0.1 Hz was applied to each time series. No spatial smoothing was performed. After the above preprocessing, in order to match the voxel size of the thalamic parcellation result, we resampled the voxel size of functional images into isotropic 2 mm. Each cluster identified by MAPBOT was used as a "seed", and then averaged time series from each cluster (L and L and R considered together considered together) was extracted. The, partial correlation was performed between each cluster time series and the every voxel time series from the whole brain, eliminating the shared variance among all the thalamic clusters. Then one-sample *t*-tests were performed on each set of functional connectivity map across all subjects. The results were thresholded at FWE-corrected $p < 0.05$ at the voxel level (Fig. 5).

Results

We demonstrate the use of the text features to segment the human thalamus, as well as the criteria to determine the "optimal" cluster. Due to the dual-view property of NMF, we also derive the topic mapping of each thalamic cluster.

Thalamic sub-divisions

From the calculated HI and VI values, as well as the symmetric VI, the thalamus was divided into 10 different sub-regions on both the left and right sides. Considering the location and overlaps with the Morel template, the left and right clusters were matched accordingly (Fig. 2).

Clusters 1 and 2 were located in the anterior part of the thalamus, while Clusters 8 were in the dorsal part. Cluster 4 is just alone the midline which is the major medial section of the thalamus. Clusters 3, 5 were in the ventral and lateral parts of thalamus, and Clusters 6, 7 and 9 were in the posterior part (Fig. 3). We examined the overlaps between the MAPBOT parcellation masks and the 11 major nuclei (38 subnuclei) of the (downsampled) Morel template. The complete list of overlaps between our parcellation masks and the Morel template are given in the Tables S1–2.

Cluster 1 includes the anterior ventral nucleus (AV), ventral anterior nucleus (VA), central medial nucleus (CeM),. Cluster 2 encompasses the lateral dorsal nucleus (LD), the ventral lateral posterior nucleus (VLp) and the dorsal division (VLpd). Cluster 3 encompasses the ventral medial nucleus (VM), ventral lateral anterior nucleus (VLa), ventral lateral posterior nucleus, paralamellar division (VLpv), ventral posterior nucleus. Cluster 4 is overlapped with the major part of mediodorsal nucleus, habenular nucleus (Hb). Cluster 5 overlaps the anterior pulvinar (PuA), ventral posterior nucleus (VP), posterior nucleus of hypothalamus (PO), suprageniculate nucleus(SG). Cluster 6 encompasses the lateral pulvinar (PuL). Cluster 7 is largely overlapped with the limitans nucleus (Li), medial pulvinar (PuM). Cluster 8 encompasses at the lateral posterior nucleus (LP), lateral dorsal nucleus (LD), and part of the ventral lateral posterior nucleus (paralamellar division; VLpl). Cluster 9 encompasses the lateral pulvinar (PuL), lateral geniculate nucleus (LGN), and medial geniculate nucleus (MGN). Cluster 10 overlaps the red nucleus (RN) subparafascicular nucleus (sPf) (Fig. 4)

While we note that there were some slight differences between the left and right thalamus, the majority of the parcels derived here showed large overlaps with the Morel atlas. Overall, the bilateral thalamic subdivisions have shown a notable degree of similarity between the results and the organization of major nuclei from Morel's template.

The topic mapping from the dual view of the NMF

The dual-view of decomposition of the NMF enables us to have a set of grouped terms associated with the spatial Clusters of the thalamic sub-division. The related topics are examined on the left (L) and right (R), respectively. In general, the L-R pairs of Clusters were related to similar topics across 4 topic sets (50, 100, 200, 400), which are shown in Tables 1, 2 (NB: as each topic includes around 30 terms, for brevity only the first three words are presented to represent each). As the tables shown, Cluster 1 is significantly related with social empathy, decision making (Tables 1, 2). The functional specificity of Cluster 2 might be interpreted as related to language, memory and cognitive performance, such as "semantic words word", "task performance cognitive". Cluster 3 was found to be associated with the motor related topics, such as "motor movement movements", "motor planning execution", and "stimulation somatosensory tms". Cluster 4 was found to be significantly associated with the default mode network. Cluster 5 is significantly associated with the "motor sensory areas", "action actions observation" and "auditory visual sensory," "speech auditory production." Cluster 6 is found to be significantly associated with "eye spatial gaze," "eye eyes movement" on left and "learning training performance" on the right. Cluster 7 on the left is significantly associated with Alzheimer's disease, mild cognitive

impairment, while on the right, Cluster 7 is related to “emotional faces facial,” “faces face facial” and “volume grey structural.” Cluster 8 was found related to “Schizophrenia patients” and “age adults aging,” “training practice trained.” Cluster 9 was significantly associated with “memory encoding and retrieval,” “object recognition,” and “visual perception.” Cluster 3 had a higher tendency toward “reward monetary anticipation,” and similar topics such as “inhibition response inhibitory.” To further illustrate the general idea of the topic mapping, topic-mapping result from the topic set of 50 is used to demonstrate the relation between topics and clusters (Figs. 5 and 6).

Decoding of thalamic sub-divisions

The behavior topic for each cluster was examined through both forward and inverse inference. In general, the forward inference indicates, given a topic, which cluster is most strongly associated with it, while the inverse inference determines, given the cluster coordinates, what specific topic the cluster corresponds to.

First, we examined the most over-represented topics across all thalamic clusters, which exhibited such varied topics as motor, language, emotion, morals, and speech. To further look into the details of each cluster, we investigated the preferential associations of each thalamic sub-division toward particular topics (Table S3, S4). The posthoc topic mapping is highly overlapped the topic mapping results from dual-view, but more general and less specific than dual-view topic mapping. In a brief summary, Cluster 1 is associated with reward and social empathy related topics; Cluster 2 is involved with working memory and pain related topics. Cluster 3, 5, 8 are all associated with motor related function, though each of them are from different perspective, Cluster 4 consists of several different kinds of topics, such as, pain, emotion, reward, arousal, task preparatory. Cluster 6 is associated only with visual related topics, Cluster 7 is highly related with Alzheimer’s and mild cognitive dementia. Cluster 9 is related with memory dependent topics. Cluster 10 is related with inhibition, reward, and motor planning.

Resting state functional connectivity

The resting-state functional connectivity (RSFC) maps of the thalamic sub-divisions are shown in Fig. 7. All 10 RSFC maps revealed that most sub-divisions within the thalamus showed strongly positive correlations with distinct cortical regions (Table 3). Cluster 1 appears to be functionally associated with the anterior cingulate cortex; Cluster 2 is largely associated with frontal lobe, including inferior, middle and superior frontal gyrus. Cluster 3 demonstrates functional connectivity to the precentral gyrus and supplementary motor area, as well as the cerebellar structures; Cluster 4 is functional connected to anterior and middle cingulate cortex, precuneus, and middle frontal gyrus; Cluster 5 is highly related with motor function related areas, like precentral gyrus, SMA, and superior temporal gyrus, insula; Cluster 6 is connected with the cuneus, posterior cingulate, and lingual gyrus; Cluster 7 is associated with the middle temporal gyrus and posterior cingulate gyrus; Cluster 8 connects to the precuneus, inferior parietal lobule, insula, and postcentral gyrus; Cluster 9 shows connection with middle orbital gyrus, superior and middle temporal gyrus, inferior and middle frontal gyrus; Cluster 10 is related to the claustrum, hippocampus, amygdala, and fusiform gyrus.

Discussion

We examined the thalamic organization based upon the textual terms and demonstrated an alternative meta-analysis approach called MAPBOT to characterize the sub-divisions of the thalamus (Fig. 3). The NMF method is the core of our method, which is widely used to cluster documents and here adapted to parcellate the thalamus based on voxel- to-term relations. The clustering results and evaluation of behavioral inference from parceled clusters showed differential thalamocortical relation among different thalamic sub-divisions (Tables 1, 2, S3–4, Figs. 5 and 6). The resting-state functional connectivity map was estimated for each cluster (Fig. 7, Table 3).

Thalamic sub-divisions: topic mapping and relations to preexisting templates

The thalamus is known to be related with several cognitive functions, such as motor, emotion, and memory processing (Jones et al., 2007). The sub-divisions of the thalamus have been found to be specifically connected to distinct cortical regions (Jones et al., 2007). Previous functional neuroimaging studies have provided examples of parcellating the thalamus into 7, 9, or 10 clusters, based on several different classifications (Behrens et al., 2003; Ji et al., 2016; Kim et al., 2013; Fan et al., 2015). In addition, recent work suggests a more complicated organization of the thalamic nuclei and that higher order nuclei might be involved within multiple functions (Yuan et al., 2016). In the current study, we delineated the thalamus based upon a metaanalysis and expanded on the results of previous studies which were mainly based on resting-state functional connectivity or white matter connection.

In the present study, the final parcellation had 10 sub-regions in each hemisphere, as the best solution from a wide range of parcel numbers investigated (in the range of 2–20 sub-regions).

Cluster 1 is at the anterior and dorsal region of the thalamus, encompasses the anterior nucleus (AN), ventral anterior nucleus (VA), and the central medial nucleus (CeM). The VA consists of two major portions, one is called the magnocellular (VAmc), and the other one is called the parvocellular (VApc). Interestingly, unlike other ventral nuclei, which project to specific cortical regions, the VA has a dense fiber network between itself and several thalamic nuclei (mostly VApc), and it widely projects to the frontal lobe cortex (mostly from the VAmc) as well as being orbitofrontal afferent to the VAmc nucleus (Carmel et al., 1970; Scheibel et al., 1966). Animal studies have indicated that the VAmc also plays an important role in executive function with cognitive context and emotional processing, since it is densely projection from anterior cingulate areas, dorsolateral prefrontal areas and the basal ganglia, especially the global pallidus and the substantia nigra reticutata (Xiao and Barbas et al., 2002, 2004; McFarland et al., 2002). Although it is not completely overlapped with the anatomical connections based on animal studies, the human-derived FC demonstrates the significant connection between Cluster 1 and the anterior cingulate gyrus, and basal ganglia. The topic mappings of the Cluster 1 are found to be strongly associated with reward and emotion related topics. Moreover, Cluster 1 also encompasses the major part of the anterior nuclei (AN), the medial portion of AN has reciprocal connections with the anterior cingulate and orbitomedial prefrontal cortices, and relays information to medial frontal lobe. It might be also involved in emotional and executive functions through the hippocampal-prefrontal

interactions (Child et al., 2013). Furthermore, previous studies have also suggested that the AN with emotional experience might modify and enhance hippocampal processes involved in memory (Child et al., 2013; Papez et al., 1937; Vertes et al., 2001). However, here the topic mapping itself has not directly shown any memory related function.

Cluster 2, dorsal-rostral to Cluster 1, which includes ventral lateral posterior nucleus dorsal division (Vlpd), the lateral dorsal nucleus (LD), and a small portion of the anterior ventral nucleus (AV). VLp is suggested as the principal cerebellar thalamic relay nucleus due to its afferent connection with cerebellum, and reciprocal connection with motor areas (Sakai et al., 2000). Functional connectivity studies have also shown Cluster 2 to be connected to several frontal and cerebellum structures, but not with any motor related areas. Moreover, VLp is regarded as an integral part of both the cerebello-thalamocortical and the basal ganglia-thalamocotical circuit (Danos et al., 2002; McFarland and Haber et al., 2002). On the other hand, previous animal studies have shown that the LD is connected with the posterior cingulate and parietal cortices reciprocally, as well as the hippocampal complex (Van Groen and Wyss et al., 1992; Mizumori et al., 1993; Thompson and Robertson et al., 1987). Moreover, it has also been indicated that LD integrates motivation and/or attention with sensory processes (Taber et al., 2004). The topic mapping results indicate Cluster 2 is related with the following: sematic words, working memory and task performance. Although several studies (Van Groen et al., 2002; Mizumori et al., 1994) have indicated LD's involvement in spatial memory, it is possible that Cluster 2 might have a much broader spectrum of functions due to its wide-spread anatomical and functional connections.

Cluster 3 has the largest volume in the thalamic parcellation within this present study. It encompasses the ventral medial nucleus (VM), ventral lateral anterior nucleus (VLa), ventral lateral posterior nucleus, and the paralamellar division (VLpv). The VM lies between the VL and VP as a transition zone, and it has not always been considered to be an independent nucleus. It has been suggested to be related to motor function and the gustatory relay (Buee et al., 1986; Craig et al., 2014). The VL has long been considered as the "motor thalamus", which consisted of 3 distinct clusters. VLa generally projects to the premotor cortex and SMA, while VLp connects to the primary motor cortex, globus pallidus and cerebellum (Asanuma et al., 1983; Jones et al., 1979; Stepniewska et al., 1994). Cluster 3 is also shown to be functionally connected with precentral gyrus and SMA. The topic mapping results here are also strongly aligned with these anatomical and functional connections, specifically the topic of "motor movement movements".

Cluster 4, located at the medial part of the thalamus, encompasses the major part of mediodorsal nucleus (MD), central laberal (CL) and habenular nucleus (Hb). The MD might be the most well studied thalamic nuclei. It has major reciprocal connections with the prefrontal, anterior cingulate cortex, temporal lobe and supplementary motor cortex in primates (Goldman-Rakic et al., 1985; Russchen et al., 1987; Ray and Price, 1993; Siwek and Pandya, 1991; Vogt, 1979). It also receives afferent information from subcortical regions, such as the amygdala, the basal forebrain, and the cerebellum (Hreib et al., 1988; Aggleton et al., 1980; Aggleton and Mishkin et al., 1984; Price et al., 1986). The MD is a key to multiple brain functions, such as cognition, emotion, pain, memory, sleep/waking circuits, and motivation (Haber and McFarland et al., 2001; Sherman and Guillery, 2013).

Furthermore, left and right dual-view topic mapping have shown its relation with the default mode network (DMN), which is supported by the association of MD with DMN by large groups of functional studies (Zhang et al., 2010), though it might be more closely related with posterior DMN (Yuan et al., 2016). In the current study the estimated RSFC of Cluster 4 did not reflect a standard DMN pattern, nor did any other clusters. It is possible that the DMN does not merely connect to one single thalamic nucleus but several. However, the post-hoc decoding is highly related to topics such as pain, autonomic arousal, emotion, and reward. Since the meta-analysis underlying the parcellation has originated from published studies, certain function related studies might be published significantly more than other studies (Yarkoni et al., 2011), for structures like MD, which has substantially related function (Golden et al., 2016; Mitchell et al., 2015; Metzger et al., 2010). It might be impossible for topic mapping to list all possible functions, but only to certain highly represented research topics.

Cluster 5 consisted of the PuA, the anterior and medial and inferior VP, posterior nuclus (PO), suprageniculate nucleus (SG) as well as the MGN. Previous animal studies on macaques suggested that both the anterior pulvinar and ventral posterior superior nuclei are connected to the area 2 of somatosensory cortex (Cusick et al., 1990; Pons et al., 1985; Sherman and Guillery, 2013). The ventral posterior inferior nucleus (VPI) is essentially related to the sensory motor system, having been shown in primate studies to have afferent connection from the spinothalamic system and efferent connections to the second somatosensory area (Friedman and Murray, 1986; Stevens et al., 1993) and to a lesser extent S1 (Stevens et al., 1993). The resting-state functional connectivity map of Cluster 5 also indicate its connection with motor related areas, such as precentral gyrus, postcentral gyrus, Moreover, Cluster 5 is functionally connected with the auditory cortex, such as the superior temporal gyrus, which might be due to its small portion of MGN and the major part of SG. The MGN is known to be the major auditory functional relay (Winer et al., 1992). The SG has been found to be involved with auditory pathways, and response with auditory, somatic, and noxious stimulations (Kobler et al., 1987; Benedek et al., 1997). Thus, topic mapping from left and right of Cluster 5 in general are aligned as related with motor and auditory functions.

In most cases, the human pulvinar can be parceled into four parts (Morel et al., 1997), anterior (PuA), lateral (PuL), medial (PuM), and inferior (PuI). Other than PuA (part of Cluster 5, Cluster 8) which is more functionally related to somatosensory (Benarroch et al., 2015a, 2015b), Clusters 6, 7, and 9 parcels the rest of the pulvinar (PuI, PuL, PuM). Cluster 6 overlaps with the lateral portion of PuM and PuL. Cluster 7 is mainly associated with the medial portion of PuM. Cluster 9 overlaps with the inferior portion of PuL and PuM, and the whole PuI.

The dorsal PuL, as a part of **Cluster 6**, is known to connect with the visual cortex, dorsolateral prefrontal cortex, inferior parietal lobule, and superior temporal gyrus, where both those cortical regions and dorsal PuL have been long implicated as playing roles in spatial or visual attention (Mesulam et al., 1990; LaBerge et al., 1990; Robinson et al., 1993). However, RSFC of Cluster 6 is matched with the majority of the anatomical connections that described above, such as the visual cortex, parietal and posterior cingulate

areas, but not the frontal areas. On the other hand, the left and right topic mappings of Cluster 6 are mainly about “eye spatial gaze”, “object category”, “learning training performance”. The reason why the function of learning is also related to the Cluster 6 might be because that learning is highly associated with the attention mechanism. Our post-decoding of the Cluster 6 is related with topics which are well aligned with previous studies, such as “mental rotation spatial”, “space spatial location”, and “eye spatial gaze”.

Cluster 9 consisted of the inferior and lateral pulvinar, LGN. These regions project to main visual cortex regions (V1, V2, and V4), and have always been considered as the key role in the ventral stream of vision which is involved in the identification of objects, such as faces. The resting-state functional connectivity map also shows that Cluster 9 is correlated with ventral portion of visual cortex, middle orbital gyrus and cerebellum. Previous functional studies suggested that the lateral pulvinar is related with both memory and visual selection (Rotshtein et al., 2011; Johnson and Ojemann et al., 2000). For example, by electrically stimulating the lateral pulvinar, patients might develop anomia, by which they are unable to recall the names of daily objects (Ojemann et al., 1975). Our results from both the dual-view-derived and post-hoc topic mapping correspond well to these previous studies, which indicate the association with memory retrieval and object recognition. Noticeably, as the RSFC has shown, both Cluster 6 (Fig. 7F) and Cluster 9 (Fig. 7I) are connected to the visual cortex. But Cluster 6 is strongly connected to the upper visual field, and Cluster 9 connects to the lower visual field. This result might indicate two visual field maps within the pulvinar.

Cluster 7 encompasses the PuM. Previous histochemical studies on primates have provided insight into the cortical topography of efferent PuM projections. The central/lateral portion of PuM has quite widespread connection with the dorsal-lateral and orbital prefrontal cortex, as well as insula, posterior parietal areas, and posterior cingulate cortex, and the medial part of PuM has much sparser connection with prefrontal cortex, with a higher density of connections to the temporal pole, superior temporal gyrus, anterior cingulate and amygdala (Romanski et al., 1997; Jones and Burton, 1976; Aggleton et al., 1980; Benarroch, 2015a, 2015b). The functional connectivity of Cluster 7 hasshown limited connections with cortical areas, such as the middle temporal gyrus, posterior cingulate cortex. Both the left and right topic mapping and post-hoc decoding indicated that the right Cluster 7 is highly associated with Alzheimer’s disease and mild cognitive impairment. But the pulvinar is not the only thalamic nuclei that are related to Alzheimer’s disease (Braak et al., 1991; De Jong et al., 2008). It is possible that Cluster 7 is more highly associated with those related diseases than other Clusters in the collected studies, which might be due to a systematic bias.

Cluster 8 is comprised of the lateral posterior nucleus (LP), and a part of lateral dorsal nucleus (LD), ventral posterior lateral nucleus (paralamellar division; VPlp), as well as small portion of PuA. In rats, the LP is traditionally considered to be a part of the LP-pulvinar complex, which transfers visual information to the superior colliculus and striatum, and plays an important role in selecting salient visual targets and directing eye movements towards these targets (Berson et al., 1978; Kelly et al., 2003). However, recent studies on the projections from three subdivisions of the LP suggested that LP neurons have relatively widespread cortical projections to the temporal area, the dorsal perirhinal, primary somatosensory, and posterior parietal cortices, where the LP might be involved within

visuospatial processing, learning and memory (Kamishina et al., 2009; Nakamura et al., 2015). As mentioned above, PuA also connects to the somatosensory cortex. The RSFC of Cluster 8 is aligned with these anatomical studies that Cluster 8 is connected with wide spread of regions, such as postcentral gyrus, cingulate cortex, insula, inferior temporal lobule, occipital gyrus. Consistent with much of these studies, the post-hoc decoding of Cluster 8 here indicates that it is strongly associated with the motor functions. But the left and right topic mapping indicates it is related with training and practice, aging, schizophrenia. Considering LP is involved in function as selecting salient visual targets, it might be reasonable to associate topic of training with Cluster 8. Actually, there are several studies found reduced grey matter volume at pulvinar, ventral posterior nucleus in schizophrenia patients (Danos et al., 2002; Dorph-Petersen et al., 2017). Although it is highly possible that thalmocortical disfunction might contribute toschizophrenia, more studies are needed to narrow down the bio-marker to a specific thalamic sub-division.

Cluster 10 is mostly comprised of the red nucleus (RN), parafascicular nucleus (Pf), sub-parafascicular nucleus (sPf) and subthalamic nucleus (STh). The RN is believed to be connected to the motor cortex and may be involved in motor activity, though some researchers have postulated that it might be integrated with other higher-level functions, such as sensory processing (Massion et la, 1988; Gruber et al., 2010). The Pf projects massively to the basal ganglia, and efferent to the caudate nucleus, the rostral putamen the nucleus accumbens and the olfactory tubercle. Pf and sPf are closely related. In addition, the Pf- sPf complex projects to the hypothalamus, the substantia innominata, the peripeduncular nucleus, the amygdala as well as several brain stem structures (Sadikot et al., 1992). sPF also receives direct and indirect inputs of the spinal cord, as well as acoustic, somatosensory, and nociceptive inputs, which may target sub-divisions of the sPF accordingly (Ju et al., 1987; LeDoux et al., 1987; Nahin et al., 1988; Yasui et al., 1990). Moreover, the STh has been traditionally considered as a part of motor control processing (along with the basal ganglia), as well as a therapeutic target in the Parkinson's disease (Fasano et al., 2012). But increasing evidence from recent studies have indicated that a direct dopaminergic connection between the STh and substantia nigra strongly supports the STh's role of encoding reward-related information, or preference for reward (Darbaky et al., 2005; Lardeux et al., 2009; Espinosa-Parrilla et al., 2013). According to current parcellation method, these three structures were grouped together purely based on their functional similarity. Although no spatial information contributed to the clustering method, the close geographic distances between these nuclei might be a latent factor that led to this result. As agreed with anatomical studies, functional connectivity results of Cluster 10 shows its connection with amygdala, hippocampus. Given the functional and anatomical connection of Cluster 10, its topic mapping indicates most related functions, which in general pointed to reward and emotion topics, such as "reward anticipation monetary", "emotion regulation emotional", or "anxiety threat disorders".

In summary, the parcellation results group thalamic sub-divisions with similar function or cortical connections together, and follow similar nucleus distribution and "geography" to the well-known Morel thalamus template. Mismatches between these two are inevitable due to the intrinsic methods and different perspectives. Some failed predication might be due to insufficient data or limitations of this method, which will be discussed in the next section.

MAPBOT methodology: advantages and limitations

In general, the MACM based parcellation may require considerable numbers of experiments that report certain voxels or ROIs. However, this present text-based parcellation approach is much less affected by lower numbers of reports. In addition, the NMF is also relatively easy to implement, and well suited for sparse matrix applications, which is generally the case for the text feature matrix. Therefore, MAPBOT can be a complementary method to the existing MACM method. Future studies could also combine both spatial and text information together to conduct a comprehensive clustering.

Nevertheless, the NMF methodology does contain limitations. First, there does not exist an optimal means for selecting an initial matrix. Although the normalized weight- and SVD-derived initial matrix was used to define the “best” initial matrix here, the time it takes to converge is still longer than for any clustering method. Here, the approach utilized for convergence was the most popular method in the field, multiplicate updates (MU), which was proposed by Lee and Seung (1999, 2001). To enhance its rate of convergence, we also reinitialized the zero of matrix U when partial derivatives turn negative. Finally, for any clustering analysis, the stability of the method is an important measure. For example, Thirion et al. (2014) reviewed four of the most widely used clustering methods, including k-means, ward, spectral, and geometric, and each only demonstrated moderate reliability. Several studies (Xu et al., 2003; Gillis et al., 2014) suggested NMF could have a better reliability and robustness than other clustering method, such as spectral clustering. Here, we note that results did not seem highly sensitive to individual parameters during the testing and development of the method. Moreover, by combing VT matrix across different research radius ranging from 4 mm to 12 mm, current method eliminates the bias derived from certain the search range. Although several successive implementations of NMF on text mining have been reported, future improvements to the method may include using an ensemble clustering method to further improve reliability.

The post-hoc topic mapping is a similar procedure to what previous studies have done for making inverse inferences. In the present study, however, due to unknown the intrinsic frequency or prior possibility of each term, inverse inference loses some of its power to address the functional or behavioral inferences within given activation regions (Wager et al., 2016). Even so, it does bolster the functional and behavioral relation with given brain regions.

Another limitation of the method of this study is that it does not take into account effect sizes or study sizes when combining data. In large part, this is due to the fact that many studies do not often report such information (Chen et al., 2016). At present, most meta-analysis methods combine data from studies with vastly different group sizes, and ignore the physical size (where applicable) of effects. Future work should consider including such information as additional weights in the VT matrices once it is available within the database.

Conclusions

MAPBOT (Meta-Analytic Parcellation Based On Text) was introduced and implemented. This data-driven method provided a tool to understand and synthesize meta-data, providing

insights using the text mining of documents referred to as topic mapping within the neuroscience field. Here, we presented a parcellation of the thalamus based on text features of a large number of scientific papers, as well as topic mappings associated with each cluster. Our results suggest that this text oriented parcellation can be a very useful approach to parcellate sub-regions within the brain. Our algorithm has the potential to lead to a more robust and ubiquitous understanding of any brain region by pooling vast amount of information published within the scientific literature and performing MAPBOT on those papers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This research was supported by NIH 5R01NS049176 (BBB).

References

Aggleton JP, Burton MJ, Passingham RE, 1980 Cortical and subcortical afferents to the amygdala of the rhesus monkey (Macaca mulatta). Brain Res 190, 347–368. [PubMed: 6768425]

Aggleton JP, Mishkin M, 1984 Projections of the amygdala to the thalamus in the cynomolgus monkey. J. Comp Neurol 222, 56–68. [PubMed: 6321564]

Andreasen NC, Arndt S, Swayze V, Cizadlo T, Flaum M, O'Leary D, Ehrhardt JC, Yuh WT, 1994 Thalamic abnormalities in schizophrenia visualized through magnetic resonance image averaging. Science 266, 294–298. [PubMed: 7939669]

Asanuma C, Thach WT, Jones EG, 1983 Cytoarchitectonic delineation ofthe ventral lateral thalamic region in the monkey. Brain Res 286, 219–235. [PubMed: 6850357]

Ashburner J, 2007 A fast diffeomorphic image registration algorithm. Neuroimage 38, 95–113. [PubMed: 17761438]

Barron DS, Eickhoff SB, Clos M, Fox PT, 2015 Human pulvinar functional organization and connectivity. Hum. Brain Mapp. 36, 2417–2431. [PubMed: 25821061]

Beam E, Appelbaum LG, Jack J, Moody J, Huettel SA, 2014 Mapping the semantic structure of cognitive neuroscience. J. Cogn. Neurosci. 26, 1949–1965. [PubMed: 24666126]

Behrens TEJ, Johansen-Berg H, Woolrich MW, Smith SM, Wheeler-Kingshott CAM, Boulby PA, Barker GJ, Sillery EL, Sheehan K, Ciccarelli O, Thompson AJ, Brady JM, Matthews PM, 2003 Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat. Neurosci. 6, 750–757. [PubMed: 12808459]

Benarroch EE, 2015a Pulvinar Associative role in cortical function and clinical correlations. Neurology 84 (7), 738–747. [PubMed: 25609762]

Benarroch EE, 2015b Pulvinar: associative role in cortical function and clinical correlations. Neurology 84 (7), 738–747. [PubMed: 25609762]

Benedek G, Pereny J, Kovacs G, Fischer-Szatmari L, Katoh YY, 1997 Visual, somatosensory, auditory and nociceptive modality properties in the feline suprageniculate nucleus. Neuroscience 78, 179–189. [PubMed: 9135099]

Berry MW, Browne M, Langville AN, Pauca VP, Plemmons RJ, 2007 Algorithms and applications for approximate nonnegative matrix factorization. Comput. Stat. dataAnal. 52, 155–173.

Berson DM, Graybiel AM, 1978 Parallel thalamic zones in the LP-pulvinar complex of the cat identified by their afferent and efferent connections. Brain Res. 147, 139–148. [PubMed: 656909]

Biswal BB, Mennes M, Zuo X-N, Gohel S, Kelly C, Smith SM, Beckmann CF, Adelstein JS, Buckner RL, Colcombe S, Dogonowski A-M, Ernst M, Fair D, Hampson M, Hoptman MJ, Hyde JS, Kiviniemi VJ, Kotter R, Li S-J, Lin C-P, Lowe MJ, Mackay C, Madden DJ, Madsen KH,

Margulies DS, Mayberg HS, McMahon K, Monk CS, Mostofsky SH, Nagel BJ, Pekar JJ, Peltier SJ, Petersen SE, Riedl V, Rombouts SARB, Rypma B, Schlaggar BL, Schmidt S, Seidler RD, Siegle GJ, Sorg C, Teng G-J, Veijola J, Villringer A, Walter M, Wang L, Weng X-C, Whitfield-Gabrieli S, Williamson P, Windischberger C, Zang Y-F, Zhang H-Y, Castellanos FX, Milham MP, 2010 Toward discovery science of human brain function. Proc. Natl. Acad. Sci. USA 107, 4734–4739. [PubMed: 20176931]

Boutsidis C, Gallopoulos E, 2008 SVD based initialization: a head start for nonnegative matrix factorization. Pattern Recognit. 41, 1350–1362.

Braak H, Braak E, 1991 Alzheimer's disease affects limbic nuclei ofthe thalamus. Acta Neuropathol. 81, 261–268. [PubMed: 1711755]

Brown MB, 1975 400: a method for combining non-independent, one-sided tests of significance. Biometrics, 987–992.

Brunet JP, Tamayo P, Golub TR, Mesirov JP, 2004 Metagenes and molecular pattern discovery using matrix factorization. Proc. Natl. Acad. Sci. USA 101, 4164–4169. [PubMed: 15016911]

Buckner RL, Petersen SE, 1996 (What does neuroimaging tell us about the role of prefrontal cortex in memory retrieval?). Elsevier, 47–55.

Buee J, Deniau JM, Chevalier G, 1986 Nigral modulation of cerebello-thalamo-cortical transmission in the ventral medial thalamic nucleus. Exp. Brain Res. 65, 241–244. [PubMed: 3803508]

Bzdok D, Langner R, Schilbach L, Jakobs O, Roski C, Caspers S, Laird AR, Fox PT, Zilles K, Eickhoff SB, 2013 Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. Neuroimage 81, 381–392. [PubMed: 23689016]

Carley KM, 1997 Network text analysis: the network position of concepts. Text. Anal. Social. Sci.: Methods Draw. Stat. Inferences texts Transcr, 79–100.

Carmel PW, 1970 Efferent projections of the ventral anterior nucleus of the thalamus in the monkey. Am. J. Anat. 128, 159–183. [PubMed: 4986961]

Cauda F, Costa T, Torta DM, Sacco K, D'Agata F, Duca S, Geminiani G, Fox PT, Vercelli A, 2012 Meta-analytic clustering of the insular cortex: characterizing the meta-analytic connectivity of the insula when involved in active tasks. Neuroimage 62, 343–355. [PubMed: 22521480]

Chang LJ, Yarkoni T, Khaw MW, Sanfey AG, 2013 Decoding the role of the insula in human cognition: functional parcellation and large-scale reverse inference. Cereb. cortex 23, 739–749. [PubMed: 22437053]

Chen G, Taylor PA, Cox RW, 2016 Is the statistic value all we should care about in neuroimaging? Neuroimage.

Child ND, Benarroch EE, 2013 Anterior nucleus of the thalamus: functional organization and clinical implications. Neurology 81, 1869–1876. [PubMed: 24142476]

Corradi-Dell'Acqua C, Tomelleri L, Bellani M, Rambaldelli G, Cerini R, Pozzi-Mucelli R, Balestrieri M, Tansella M, Brambilla P, 2012 Thalamic-insular dysconnectivity in schizophrenia: evidence from structural equation modeling. Hum. Brain Mapp. 33, 740–752. [PubMed: 21484952]

Cox RW, 1996 AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173. [PubMed: 8812068]

Craig AD, 2014 Topographically organized projection to posterior insular cortex from the posterior portion of the ventral medial nucleus in the long-tailed macaque monkey. J. Comp. Neurol. 522, 36–63. [PubMed: 23853108]

Cusick CG, Gould HJ 3rd, 1990 Connections between area 3b of the somatosensory cortex and subdivisions of the ventroposterior nuclear complex and the anterior pulvinar nucleus in squirrel monkeys. J. Comp. Neurol. 292, 83–102. [PubMed: 1690224]

Danos P, Baumann B, Bernstein HG, Stauch R, Krell D, Falkai P, Bogerts B, 2002 The ventral lateral posterior nucleus of the thalamus in schizophrenia: a postmortem study. Psychiatry Res. 114, 1–9. [PubMed: 11864805]

Darbaky Y, Baunez C, Arecchi P, Legallet E, Apicella P, 2005 Reward-related neuronal activity in the subthalamic nucleus of the monkey. Neuroreport 16, 1241–1244. [PubMed: 16012357]

Daube-Witherspoon ME, Muehllehner G, 1986 An iterative image space reconstruction algorithm suitable for volume ECT. IEEE Trans. Med. Imaging 5, 61–66. [PubMed: 18243988]

de Jong LW, van der Hiele K, Veer IM, Houwing JJ, Westendorp RG, Bollen EL, de Bruin PW, Middelkoop HA, van Buchem MA, van der Grond J, 2008 Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. Brain :J. Neurol. 131, 3277–3285.

Diesner J, Carley KM, 2004 Revealing social structure from texts. Causal Mapp. Res. Inf. Technol, 81.

Dorph-Petersen KA, Lewis DA, 2017 Postmortem structural studies of the thalamus in schizophrenia. Schizophr. Res 180, 28–35. [PubMed: 27567291]

Draganski B, Kherif F, Kloppel S, Cook PA, Alexander DC, Parker GJM,

Deichmann R,Ashburner J, Frackowiak RSJ, 2008 Evidence for segregated and integrative connectivity patterns in the human Basal Ganglia. J. Neurosci. 28, 7143–7152. [PubMed: 18614684]

Espinosa-Parrilla JF, Baunez C, Apicella P, 2013 Linking reward processing to behavioral output: motor and motivational integration in the primate subthalamic nucleus. Front. Comput. Neurosci. 7, 175. [PubMed: 24381555]

Fan Y, Nickerson LD, Li H, Ma Y, Lyu B, Miao X, Zhuo Y, Ge J, Zou Q, Gao J-H, 2015 Functional connectivity-based parcellation of the thalamus: an unsupervised clustering method and its validity investigation. Brain Connect. 5, 620–630. [PubMed: 26106821]

Fasano A, Daniele A, Albanese A, 2012 Treatment of motor and non-motor features of Parkinson's disease with deep brain stimulation. Lancet Neurol. 11, 429–442. [PubMed: 22516078]

Fox PT, Lancaster JL, 2002 Opinion: mapping context and content: the BrainMap model. Nat. Rev. Neurosci. 3, 319–321. [PubMed: 11967563]

Friedman DP, Murray EA, 1986 Thalamic connectivity ofthe second somatosensory area and neighboring somatosensory fields of the lateral sulcus of the macaque. J. Comp. Neurol. 252, 348–373. [PubMed: 3793981]

Gillis N, 2014 The why and how of nonnegative matrix factorization. Regul., Optim., Kernels, Support Vector Mach. 12, 257.

Golden EC, Graff-Radford J, Jones DT, Benarroch EE, 2016 Mediodorsal nucleus and its multiple cognitive functions. Neurology 87, 2161–2168. [PubMed: 27770073]

Goldman-Rakic PS, Porrino LJ, 1985 The primate mediodorsal (MD) nucleus and its projection to the frontal lobe. J. Comp. Neurol. 242, 535–560. [PubMed: 2418080]

Greicius MD, Flores BH, Menon V, Glover GH, Solvason HB, Kenna H, Reiss AL, Schatzberg AF, 2007 Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. Biol. Psychiatry 62, 429–437. [PubMed: 17210143]

Gruber P, Gould D, 2010 The red nucleus: past, present, andfuture. Neuroanatomy 9, 1–3.

Haber S, McFarland NR, 2001 The place of the thalamus in frontal cortical-basal ganglia circuits. Neurosci.: Rev. J. bringing Neurobiol., Neurol. Psychiatry 7, 315–324.

Hartigan JA, 1985 Statistical theory in clustering. J. Classif. 2, 63–76.

Hreib KK, Rosene DL, Moss MB, 1988 Basal forebrain efferents to the medial dorsal thalamic nucleus in the rhesus monkey. J. Comp. Neurol. 277, 365–390. [PubMed: 2461974]

Ji B, Li Z, Li K, Li L, Langley J, Shen H, Nie S, Zhang R, Hu X, 2016 Dynamic thalamus parcellation from resting-state fMRI data. Hum. Brain Mapp. 37, 954–967. [PubMed: 26706823]

Johnson MD, Ojemann GA, 2000 The role of the human thalamus in language and memory: evidence from electrophysiological studies. Brain Cogn. 42, 218–230. [PubMed: 10744921]

Jones EG, 1998 Viewpoint: the core and matrix of thalamic organization. Neuroscience 85, 331–345. [PubMed: 9622234]

Jones EG, 2001 The thalamic matrix and thalamocortical synchrony. Trends Neurosci. 24, 595–601. [PubMed: 11576674]

Jones EG, 2007 The Thalamus 1. Cambridge university press.

Jones EG, 2009 Synchrony in the interconnected circuitry of the thalamus and cerebral cortex. Ann. N. Y Acad. Sci. 1157, 10–23. [PubMed: 19351352]

Jones EG, Burton H, 1976 A projection from the medial pulvinar to the amygdala in primates. Brain Res 104, 142–147. [PubMed: 813820]

Jones EG, Wise SP, Coulter JD, 1979 Differential thalamic relationships of sensory-motor and parietal cortical fields in monkeys. J. Comp. Neurol. 183, 833–881. [PubMed: 105020]

Ju G, Melander T, Ceccatelli S, Hokfelt T, Frey P, 1987 Immunohistochemical evidence for a spinothalamic pathway co-containing cholecystokinin- and galanin-like immunoreactivities in the rat. Neuroscience 20, 439–456. [PubMed: 2438590]

Kahnt T, Chang LJ, Park SQ, Heinzle J, Haynes J-D, 2012 Connectivity-based parcellation of the human orbitofrontal cortex. J. Neurosci. 32, 6240–6250. [PubMed: 22553030]

Kamishina H, Conte WL, Patel SS, Tai RJ, Corwin JV, Reep RL, 2009 Cortical connections of the rat lateral posterior thalamic nucleus. Brain Res. 1264, 39–56. [PubMed: 19368845]

Kelly LR, Li J, Carden WB, Bickford ME, 2003 Ultrastructure and synaptic targets of tectothalamic terminals in the cat lateral posterior nucleus. J. Comp. Neurol. 464, 472–486. [PubMed: 12900918]

Kim D-J, Park B, Park H-J, 2013 Functional connectivity-based identification of subdivisions of the basal ganglia and thalamus using multilevel independent component analysis of resting state fMRI. Hum. Brain Mapp. 34, 1371–1385. [PubMed: 22331611]

Kim PM, Tidor B, 2003 Subsystem identification through dimensionality reduction of large-scale gene expression data. Genome Res. 13, 1706–1718. [PubMed: 12840046]

Kobler JB, Isbey SF, Casseday JH, 1987 Auditory pathways to the frontal cortex of the mustache bat, Pteronotus parnellii. Science 236, 824–827. [PubMed: 2437655]

Kost JT, McDermott MP, 2002 Combining dependent P-values. Stat. Probab. Lett. 60, 183–190.

Krauth A, Blanc R, Poveda A, Jeanmonod D, Morel A, Szekely G, 2010 A mean three-dimensional atlas of the human thalamus: generation from multiple histological data. Neuroimage 49, 2053–2062. [PubMed: 19853042]

LaBerge D, Buchsbaum MS, 1990 Positron emission tomographic measurements of pulvinar activity during an attention task. J. Neurosci. 10, 613–619. [PubMed: 2303863]

Laird AR, Lancaster JL, Fox PT, 2005 BrainMap: the social evolution of a human brain mapping database. Neuroinformatics 3, 65–78. [PubMed: 15897617]

Lardeux S, Pernaud R, Paleressompoulle D, Baunez C, 2009 Beyond the reward pathway: coding reward magnitude and error in the rat subthalamic nucleus. J. Neurophysiol. 102, 2526–2537. [PubMed: 19710371]

Ledoux JE, Ruggiero DA, Forest R, Stornetta R, Reis DJ, 1987 Topographic organization of convergent projections to the thalamus from the inferior colliculus and spinal cord in the rat. J. Comp. Neurol. 264, 123–146. [PubMed: 2445791]

Lee DD, Seung HS, 1999 Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791. [PubMed: 10548103]

Lee DD, Seung HS, 2001 (Algorithms for non-negative matrix factorization). In. p 556–562.

Massion J, 1988 Red nucleus: past and future. Behav. brain Res. 28, 1–8.

McFarland NR, Haber SN, 2002 Thalamic relay nuclei of the basal ganglia form both reciprocal and nonreciprocal cortical connections, linking multiple frontal cortical areas. J. Neurosci. 22, 8117–8132. [PubMed: 12223566]

Meilă M, 2003 Comparing clusterings by the variation of information. Learning theory and kernel machines. Springer, 173–187.

Mesulam MM, 1990 Large-scale neurocognitive networks and distributed processing for attention, language, and memory. Ann. Neurol. 28, 597–613. [PubMed: 2260847]

Metzger CD, Eckert U, Steiner J, Sartorius A, Buchmann JE, Stadler J, Tempelmann C, Speck O, Bogerts B, Abler B, Walter M, 2010 High field FMRI reveals thalamocortical integration of segregated cognitive and emotional processing in mediodorsal and intralaminar thalamic nuclei. Front. Neuroanat. 4, 138. [PubMed: 21088699]

Mitchell AS, Chakraborty S, 2015 What does the mediodorsal thalamus do? Front. Syst. Neurosci. 7.

Mizumori SJ, Miya DY, Ward KE, 1994 Reversible inactivation of the lateral dorsal thalamus disrupts hippocampal place representation and impairs spatial learning. Brain Res 644, 168–174. [PubMed: 8032944]

Mizumori SJ, Williams JD, 1993 Directionally selective mnemonic properties of neurons in the lateral dorsal nucleus of the thalamus of rats. J. Neurosci. 13, 4015–4028. [PubMed: 8366357]

Morel A, Magnin M, Jeanmonod D, 1997 Multiarchitectonic and stereotactic atlas of the human thalamus. J. Comp. Neurol. 387, 588–630. [PubMed: 9373015]

Nahin RL, 1988 Immunocytochemical identification of long ascending, peptidergic lumbar spinal neurons terminating in either the medial or lateral thalamus in the rat. Brain Res 443, 345–349. [PubMed: 2896057]

Nakamura H, Hioki H, Furuta T, Kaneko T, 2015 Different cortical projections from three subdivisions of the rat lateral posterior thalamic nucleus: a single-neuron tracing study with viral vectors. Eur. J. Neurosci. 41, 1294–1310. [PubMed: 25832313]

Ojemann G, 1975 Language and the thalamus: object naming and recall during and after thalamic stimulation. Brain Lang 2, 101–120. [PubMed: 1100194]

O'Muircheartaigh J, Vollmar C, Traynor C, Barker GJ, Kumari V, Symms MR, Thompson P, Duncan JS, Koepp MJ, Richardson MP, 2011 Clustering probabilistic tractograms using independent component analysis applied to the thalamus. Neuroimage 54, 2020–2032. [PubMed: 20884353]

Owen AM, 1997 The functional organization of working memory processes within human lateral frontal cortex: the contribution of functional neuroimaging. Eur. J. Neurosci. 9, 1329–1339. [PubMed: 9240390]

Papez JW, 1937 A proposed mechanism of emotion. Arch. Neurol. Psychiatry 38,725–743.

Pauca VP, Shahnaz F, Berry MW, Plemmons RJ, 2004 (Text Mining Using Non-Negative Matrix Factorizations). SIAM, 452–456.

Poeppel D, 1996 A critical review of PET studies of phonological processing. Brain Lang. 55, 317–351, (discussion 352–85). [PubMed: 8954603]

Poldrack RA, Mumford JA, Schonberg T, Kalar D, Barman B, Yarkoni T, 2012 Discovering relations between mind, brain, and mental disorders using topic mapping. PLoS Comput. Biol. 8, e1002707. [PubMed: 23071428]

Pons T, Kaas J, 1985 Connections of area 2 of somatosensory cortex with the anterior pulvinar and subdivisions of the ventroposterior complex in macaque monkeys. J. Comp. Neurol. 240, 16–36. [PubMed: 4056103]

Popken GJ, Bunney WE, Potkin SG, Jones EG, 2000 Subnucleus-specific loss of neurons in medial thalamus of schizophrenics. Proc. Natl. Acad. Sci. USA 97, 9276–9280. [PubMed: 10908653]

Price JL, 1986 Subcortical projections from the amygdaloid complex. Adv. Exp. Med. Biol. 203, 19–33. [PubMed: 3098058]

Ray JP, Price JL, 1993 The organization of projections from the mediodorsal nucleus of the thalamus to orbital and medial prefrontal cortex in macaque monkeys. J. Comp. Neurol. 337, 1–31. [PubMed: 7506270]

Riedel MC, Ray KL, Dick AS, Sutherland MT, Hernandez Z, Fox PM, Eickhoff SB, Fox PT, Laird AR, 2015 Meta-analytic connectivity and behavioral parcellation of the human cerebellum. Neuroimage 117, 327–342. [PubMed: 25998956]

Robinson DL, 1993 Functional contributions of the primate pulvinar. Progress. Brain Res. 95, 371–380.

Robinson JL, Barron DS, Kirby LA, Bottenhorn KL, Hill AC, Murphy JE, Katz JS, Salibi N, Eickhoff SB, Fox PT, 2015 Neurofunctional topography of the human hippocampus. Hum. Brain Mapp. 36, 5018–5037. [PubMed: 26350954]

Romanski L, Giguere M, Bates J, Goldman-Rakic P, 1997 Topographic organization of medial pulvinar connections with the prefrontal cortex in the rhesus monkey. J. Comp. Neurol. 379, 313–332. [PubMed: 9067827]

Rotshtein P, Soto D, Grecucci A, Geng JJ, Humphreys GW, 2011 The role of the pulvinar in resolving competition between memory and visual selection: a functional connectivity study. Neuropsychologia 49, 1544–1552. [PubMed: 21172363]

Russchen FT, Amaral DG, Price JL, 1987 The afferent input to the magnocellular division of the mediodorsal thalamic nucleus in the monkey, Macaca fascicularis. J. Comp. Neurol. 256, 175–210. [PubMed: 3549796]

Saad ZS, Reynolds RC, 2012 Suma. Neuroimage 62, 768–773. [PubMed: 21945692]

Sadikot AF, Parent A, Smith Y, Bolam JP, 1992 Efferent connections of the centromedian and parafascicular thalamic nuclei in the squirrel monkey: a light and electron microscopic study of the thalamostriatal projection in relation to striatal heterogeneity. J. Comp. Neurol. 320, 228–242. [PubMed: 1619051]

Sakai ST, Stepniewska I, Qi HX, Kaas JH, 2000 Pallidal and cerebellar afferents to pre-supplementary motor area thalamocortical neurons in the owl monkey: a multiple labeling study. J. Comp. Neurol. 417, 164–180. [PubMed: 10660895]

Scheibel ME, Scheibel AB, 1966 The organization of the ventral anterior nucleus of the thalamus. A Golgi study. Brain Res 1, 250–268. [PubMed: 4163108]

Shahnaz F, Berry MW, Pauca VP, Plemmons RJ, 2006 Document clustering using nonnegative matrix factorization. Inf. Process. Manag. 42, 373–386.

Sherman SM, 2006 Thalamus. Sch, (1.9).

Sherman SM, Guillery Rainer W., 2013 Functional Connections of Cortical Areas: a New View from the Thalamus. MIT Press.

Shipp S, 2003 The functional logic of cortico-pulvinar connections. Philos. Trans. R. Soc. B: Biol. Sci. 358, 1605.

Siwek DF, Pandya DN, 1991 Prefrontal projections to the mediodorsal nucleus of the thalamus in the rhesus monkey. J. Comp. Neurol. 312, 509–524. [PubMed: 1761739]

Stepniewska I, Preuss TM, Kaas JH, 1994 Thalamic connections of the primary motor cortex (M1) of owl monkeys. J. Comp. Neurol. 349, 558–582. [PubMed: 7532193]

Stevens RT, London SM, Apkarian AV, 1993 Spinothalamocortical projections to the secondary somatosensory cortex (SII) in squirrel monkey. Brain Res 631, 241–246. [PubMed: 7510575]

Taber KH, Wen C, Khan A, Hurley RA, 2004 The limbic thalamus. J. Neuropsychiatry Clin. Neurosci. 16, 127–132. [PubMed: 15260362]

Thirion B, Varoquaux G, Dohmatob E, Poline J-B, 2014 Which fMRI clustering gives good brain parcellations? Front. Neurosci. 8, 13. [PubMed: 24574953]

Thompson SM, Robertson RT, 1987 Organization of subcortical pathways for sensory projections to the limbic cortex. I. Subcortical projections to the medial limbic cortex in the rat. J. Comp. Neurol. 265, 175–188. [PubMed: 3320108]

Traynor C, Heckemann RA, Hammers A, O'Muircheartaigh J, Crum WR, Barker GJ, Richardson MP, 2010 Reproducibility of thalamic segmentation based on probabilistic tractography. Neuroimage 52, 69–85. [PubMed: 20398772]

Turkeltaub PE, Eickhoff SB, Laird AR, Fox M, Wiener M, Fox P, 2012 Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. Human. Brain Mapp. 33, 1–13.

van Groen T, Kadish I, Wyss JM, 2002 The role of the laterodorsal nucleus of the thalamus in spatial learning and memory in the rat. Behav. Brain Res. 136, 329–337. [PubMed: 12429394]

van Groen T, Wyss JM, 1992 Projections from the laterodorsal nucleus of the thalamus to the limbic and visual cortices in the rat. J. Comp. Neurol. 324, 427–448. [PubMed: 1383292]

Vavasis SA, 2009 On the complexity of nonnegative matrix factorization. SIAM J. Optim. 20, 1364–1377.

Vertes RP, Albo Z, Viana Di Prisco G, 2001 Theta-rhythmically firing neurons in the anterior thalamus: implications for mnemonic functions of Papez's circuit. Neuroscience 104, 619–625. [PubMed: 11440795]

Vogt BA, Rosene DL, Pandya DN, 1979 Thalamic and cortical afferents differentiate anterior from posterior cingulate cortex in the monkey. Science 204, 205–207. [PubMed: 107587]

Wager TD, Atlas LY, Botvinick MM, Chang LJ, Coghill RC, Davis KD, Iannetti GD, Poldrack RA, Shackman AJ, Yarkoni T, 2016 Pain in the ACC? Proceedings of the National Academy of Sciences, 113:E2474–E2475.

Wager TD, Lindquist MA, Nichols TE, Kober H, Van Snellenberg JX, 2009 Evaluating the consistency and specificity of neuroimaging data using meta-analysis. Neuroimage 45, S210–S221. [PubMed: 19063980]

Winer JA, Wenstrup JJ, Larue DT, 1992 Patterns of GABAergic immunoreactivity define subdivisions of the mustached bat's medial geniculate body. J. Comp. Neurol. 319, 172–190. [PubMed: 1592903]

Xiao D, Barbas H, 2002 Pathways for emotions and memory II. afferent input to the anterior thalamic nuclei from prefrontal, temporal, hypothalamic areas and the basal ganglia in the rhesus monkey. Thalamus & Related Systems 2, 33–48.

Xiao D, Barbas Helen, 2004 Circuits through prefrontal cortex, basal ganglia, and ventral anterior nucleus map pathways beyond motor control. Thalamus Relat. Syst. 2, 325–343.

Xu W, Liu X, Gong Y, 2003 (Document clustering based on non-negative matrix factorization). ACM, 267–273.

Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD, 2011 Largescale automated synthesis of human functional neuroimaging data. Nat. Methods 8, 665–670. [PubMed: 21706013]

Yasui Y, Kayahara T, Nakano K, Mizuno N, 1990 The subparafascicular thalamic nucleus of the rat receives projection fibers from the inferior colliculus and auditory cortex. Brain Res 537, 323–327. [PubMed: 1707732]

Yuan R, Di X, Taylor PA, Gohel S, Tsai YH, Biswal BB, 2016 Functional topography of the thalamocortical system in human. Brain Struct. Funct. 221, 1971–1984. [PubMed: 25924563]

Zhang D, Snyder AZ, Fox MD, Sansbury MW, Shimony JS, Raichle ME, 2008 Intrinsic functional relations between human cerebral cortex and thalamus. J. Neurophysiol. 100, 1740–1748. [PubMed: 18701759]

Zhang D, Snyder AZ, Shimony JS, Fox MD, Raichle ME, 2010 Noninvasive functional and structural connectivity mapping of the human thalamocortical system. Cereb. Cortex 20, 1187–1194. [PubMed: 19729393]

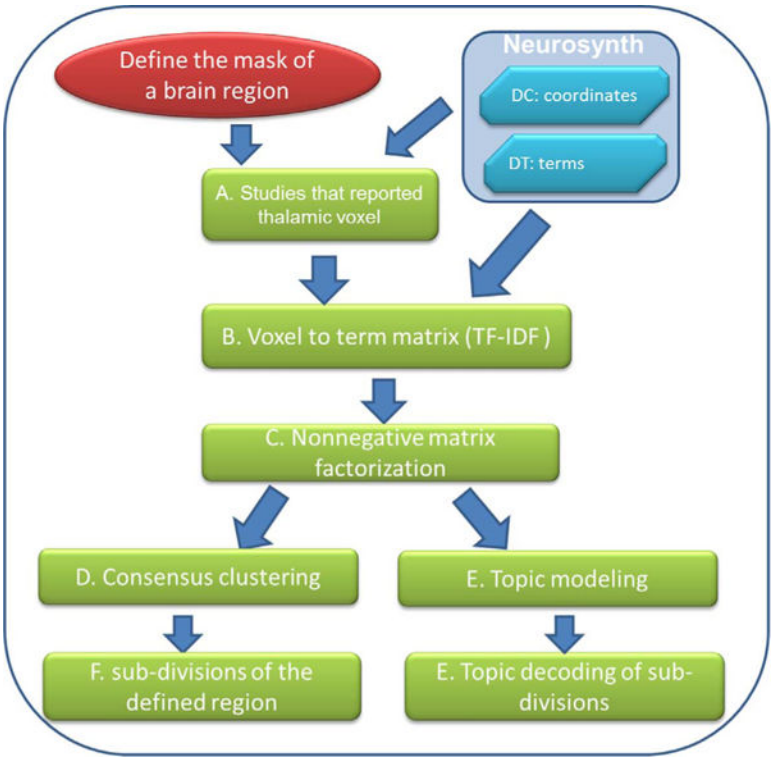

Fig. 1.

The flow chart to illustrate the steps of MAPBOT. DC is the "document to coordinate" matrix; DT is the "document to terms" matrix; *tf* is the term frequency and *idf* is the inverse document frequency (see main text). In this study, the mask of the example brain region, the thalamus, was defined by down-sampling the Morel template volume (Morel et al., 1997; Krauth et al., 2010).

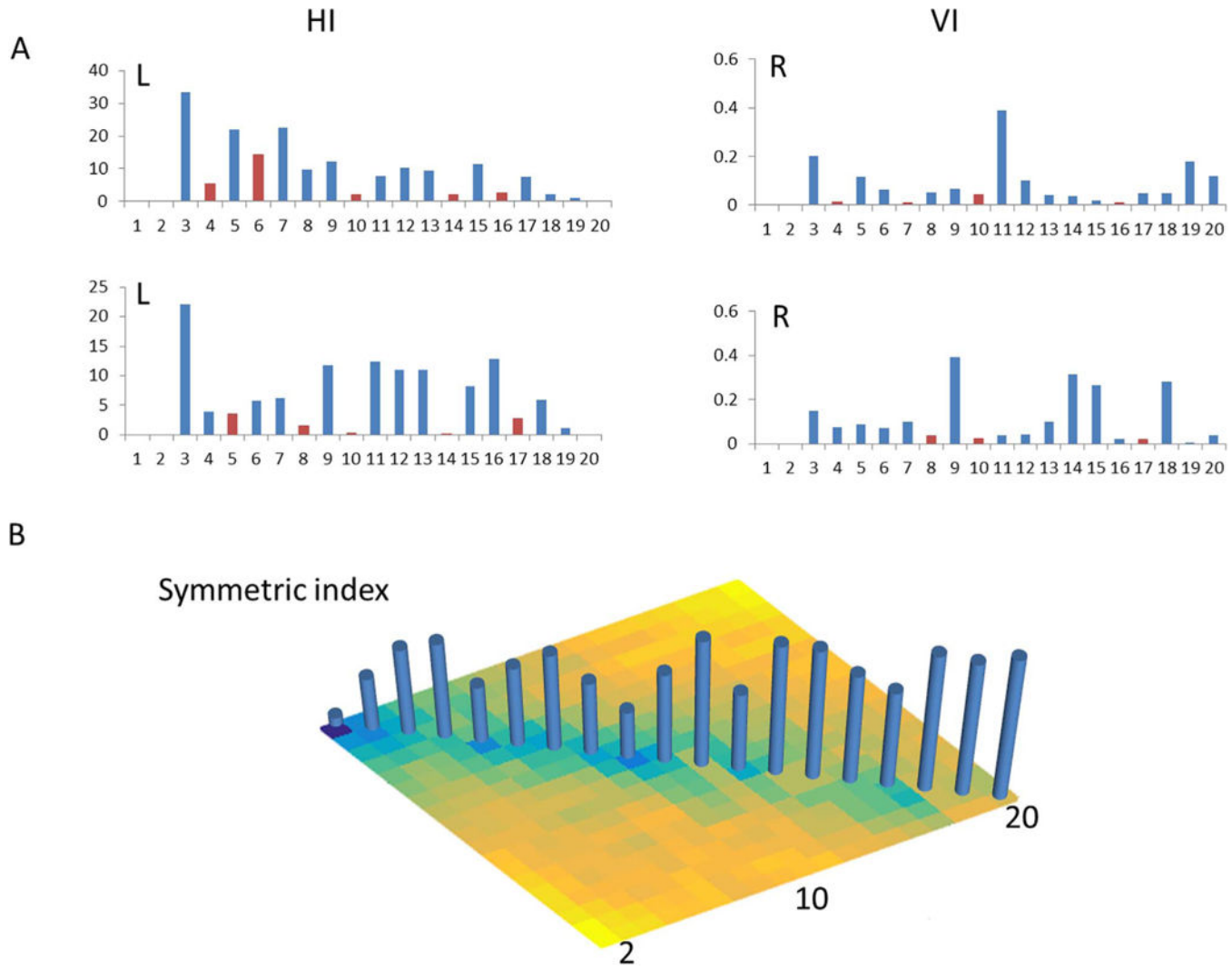

Fig. 2.

Metrics of the parcellation: the Hartigan index (HI), variance of information index (VI) and symmetric VI. The red bar indicates a potential optimal cluster solution. Shown for the applied case of the thalamus over the range of investigated factor ranks, 2<= k< = 20.

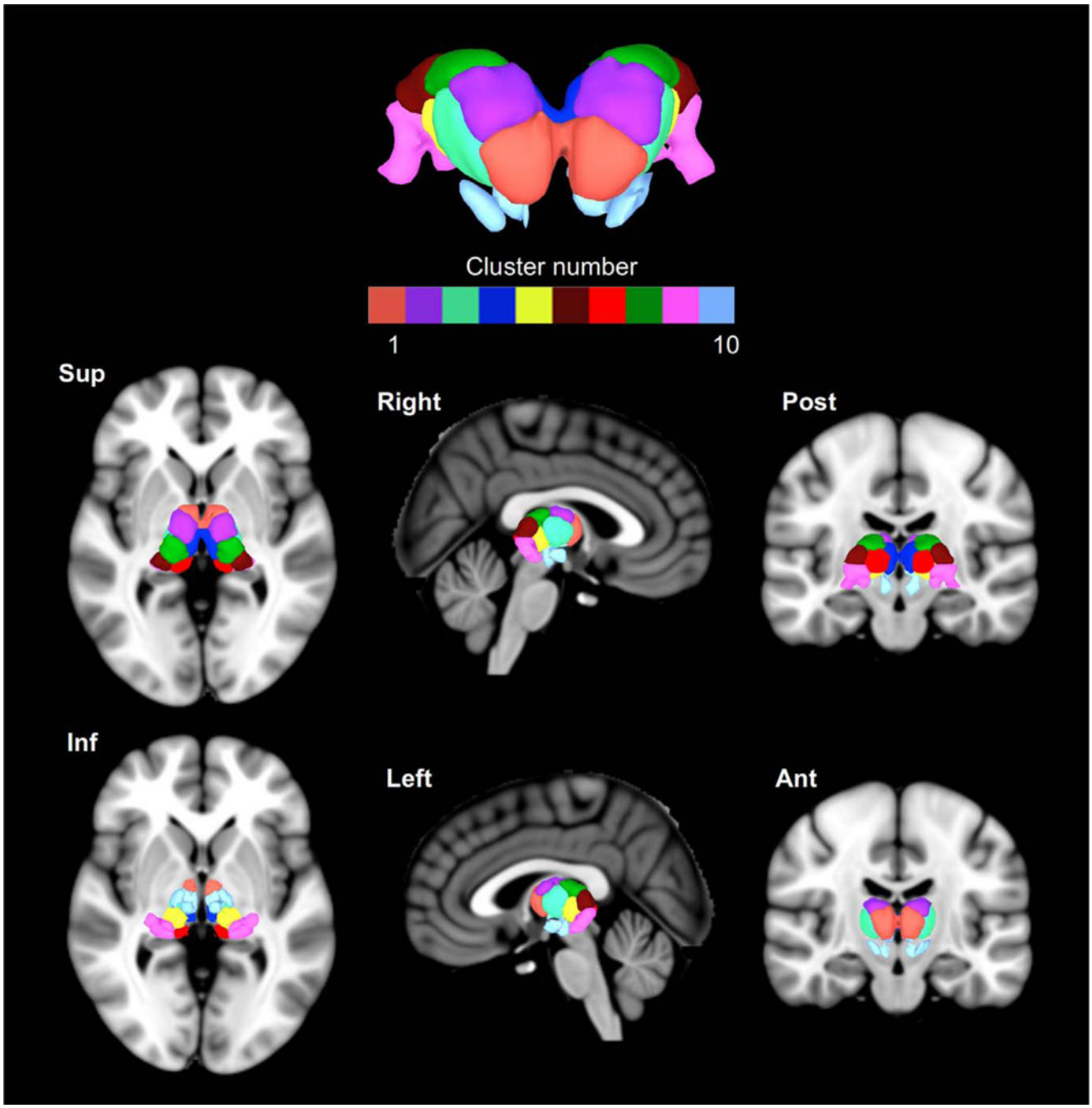

Fig. 3.

The surface view of the thalamic parcellation in SUMA (MNI space; left=left). The top panel shows a coronal (anterior) view of the parcellation, with six cardinal views below.

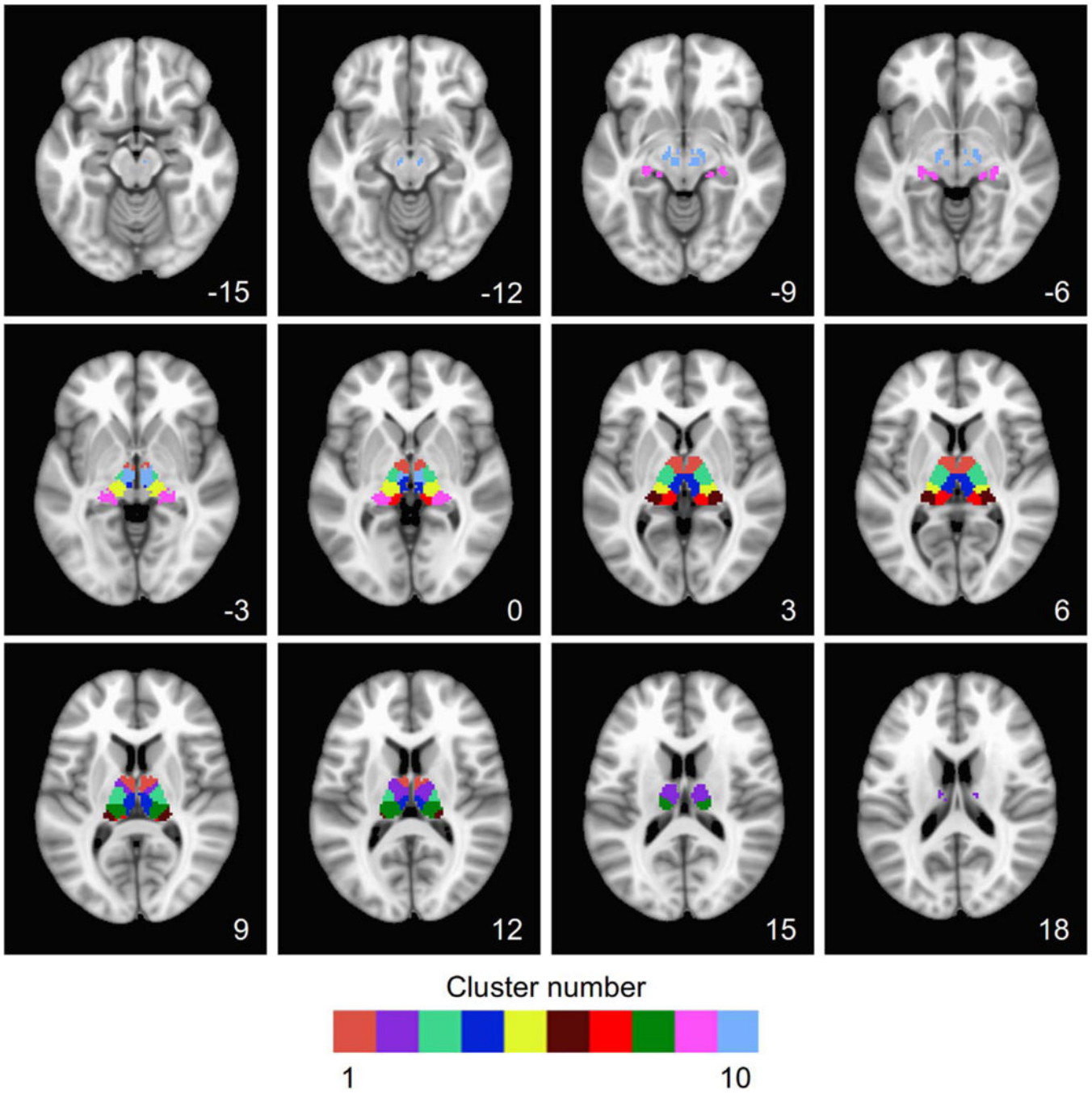

Fig. 4.

Slice-wise volumetric view of the thalamic parcellation using AFNI. Cluster numbers and colors match with Fig. 3. The z-coordinate of the axial slice (left=left) is shown in each panel.

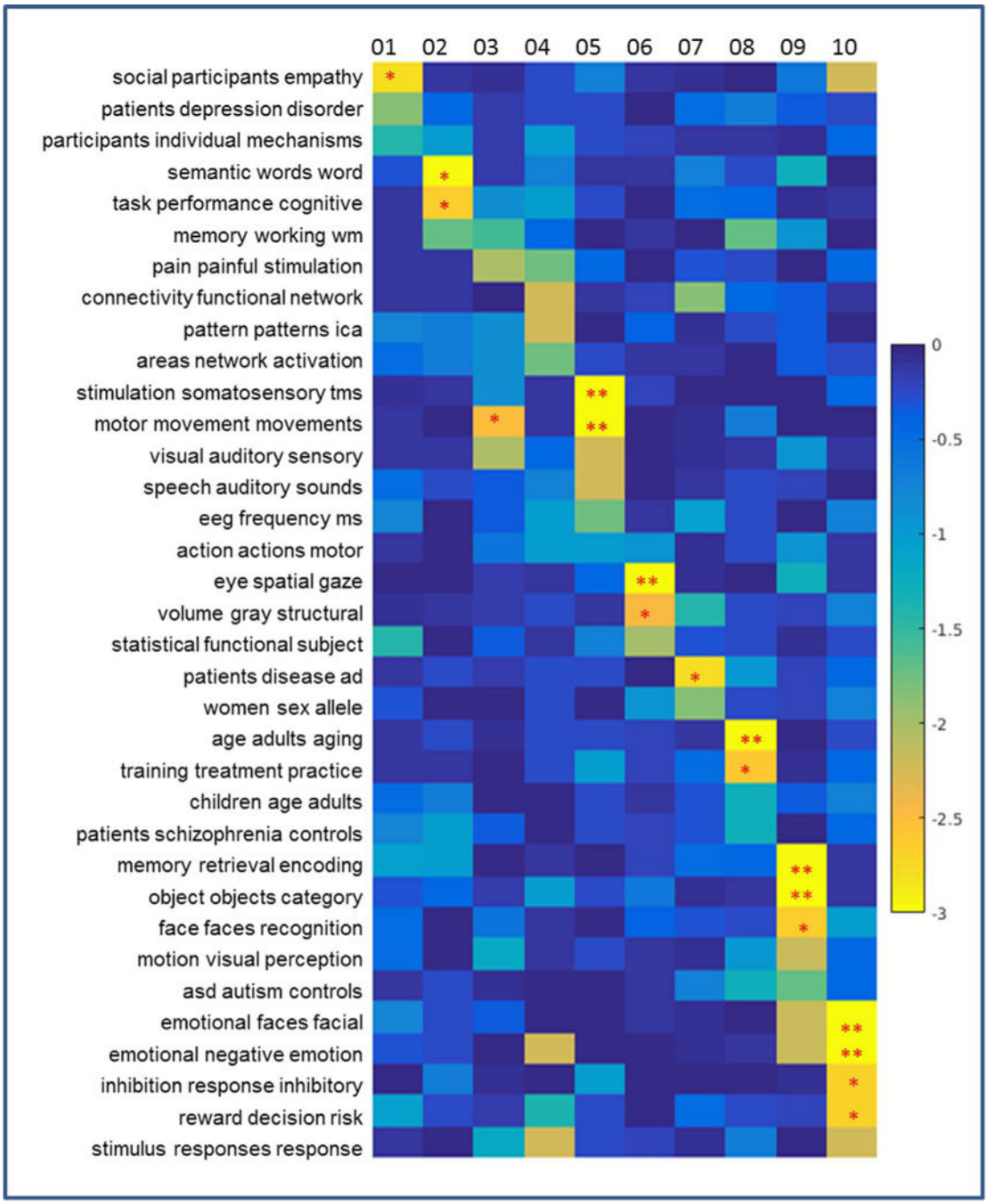

Fig. 5.

The topic mapping of Clusters 1–10 (left side) from topic set 50. This figure graphically shows relations detailed in the first column of Table 1. Each element of this map represents the significance of the relation between each topic (row) and cluster (column), shown as $\log_{10}(P)$ (thus, the lower the value, the stronger relation is; see the colorbar on the right). A single asterisk (*) indicates $P < 0.01$ (uncorrected) and two asterisks (**) indicate $P < 0.05$ (with Bonferroni correction).

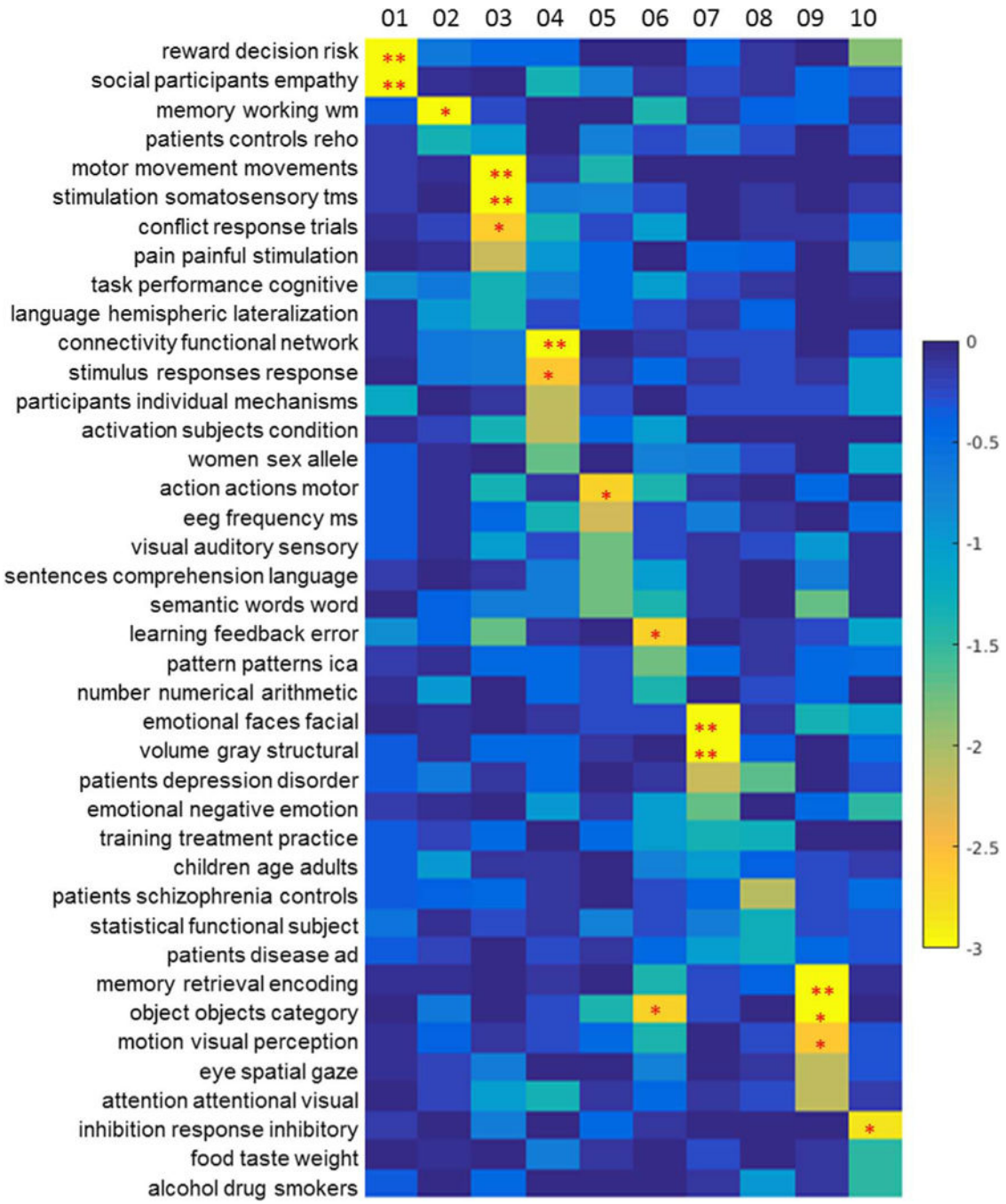

Fig. 6.

The topic mapping of Clusters 1–10 (right side) from topic set 50. This figure graphically shows relations detailed in the first column of Table 2. Each element of this map represents the significance of the relation between each topic (row) and cluster (column), shown as $\log_{10}(P)$ (thus, the lower the value, the stronger relation is; see the colorbar on the right). A single asterisk (*) indicates $P < 0.01$ (uncorrected) and two asterisks (**) indicate $P < 0.05$ (with Bonferroni correction).

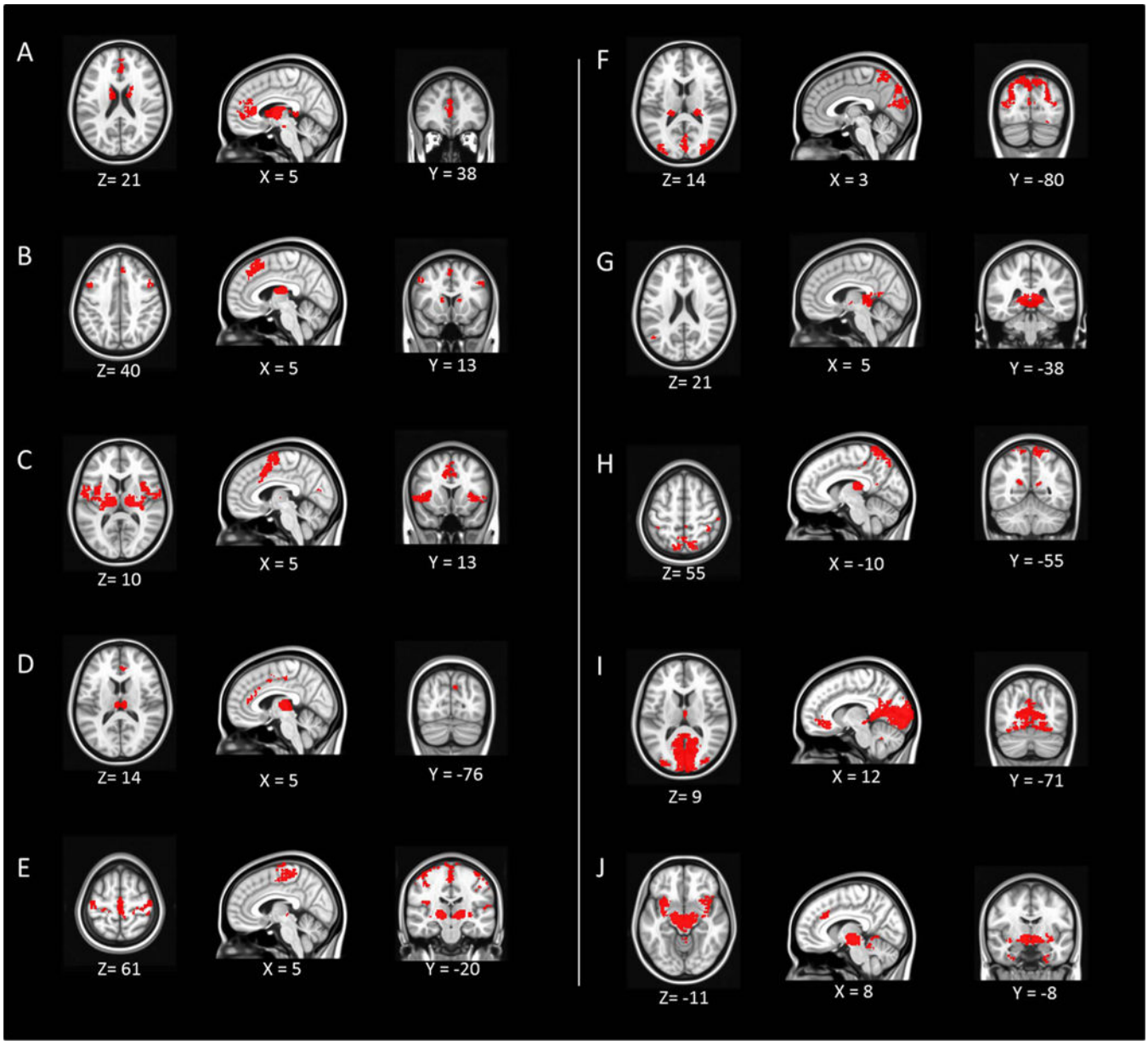

Fig. 7.

The resting-state functional connectivity of each thalamic cluster. Each panel starting from A to J corresponds to Cluster 1–10. The map is in MNI space. The x, y, z coordinates of each image is shown (left = left).

Table 1

The topic mapping of subdivisions of the left thalamus from the dual view of NMF. Highlighting shows topics that are significant at $P < 0.01$ (uncorrected), and topics in red are significant at $P < 0.05$ (Bonferroni corrected.

Left	topic 50	topic 100	topic 200	topic 400
Cluster_01	social participants empathy patients depression disorder statistical functional subject participants individual mechanisms memory retrieval encoding	social empathy participants subject individual measures arget detection stimulus depression mdd patients participants real life	empathy perspective person judgments judgment referential depression mdd patients personality trait traits negative positive affect	depression mdd depressive social cognition participants significance conclusions methodology distinct lateral dissociation session subjects scanning
Cluster_02	semantic words word task performance cognitive memory working wm memory retrieval encoding patients schizophrenia controls	cognitive control performance language verbal hemispheric switching set rule reading words language semantic words word	task load difficulty memory working verbal network networks functional processing information level trials interference task	semantic word words cognitive cognition function ability identification individual adolescents age development performance task activation
Cluster_03	motor movement movements visual auditory sensory pain painful stimulation memory working wm motion visual perception	motor movement movements connectivity effective causal pain painful stimulation object objects visual ms patients global	motor sensorimotor primary task tasks performed areas including activated pain painful stimulation activated network engaged	task tasks matching pain painful stimulation task difficulty complexity primary secondary somatosensory outcome probability expected
Cluster_04	connectivity functional network stimulus responses response pattern patterns ica emotiona negative emotion areas network activation	component ica components activation patterns pattern reward anticipation monetary semantic words word context gaze cues	correlated positively negatively cues cue preparation delay phase task component ica components activation revealed comparison	individuals participants lower autonomic rate arousal responses response evoked preparation task preparatory network default mode
Cluster_05	stimulation somatosensory tms motor movement movements visual auditory sensory speech auditory sounds eeg frequency ms	motor movement movements stimulation somatosensory tms noise echo imaging visual auditory sensory component ica components	movements motor movement tms stimulation rtms stimulation somatosensory tactile noise echo signal motor sensorimotor primary	motor finger movements motor sensory areas movement movements motor acquisition echo imaging tms stimulation motor
Cluster_06	eye spatial gaze volume gray structural statistical functional subject action actions motor women sex allele	statistical mapping maps volume gray structural eye eyes movements spatial space location action actions observation	volume gray structural eye saccades saccade maps anatomical space spatial space navigation action actions observation	volume gray structural map organization maps imaging functional resonance target targets search eye saccades saccade
Cluster_07	patients disease ad women sex allele connectivity functional network volume gray structural eeg frequency ms	disease patients ad placebo dopamine healthy connectivity functional network network default dmn treatment baseline therapy	ad disease patients attention attentional visual reho regional rs healthy subjects volunteers placebo healthy mg	ad disease mci placebo mg blind atrophy gmv disease speech auditory production healthy subjects volunteers
Cluster_08	age adults aging training treatment practice memory working wm children age adults asd autism controls	task tasks performance disease patients ad learning training performance pet flow tomography age adults children	training practice trained patients controls healthy treatment baseline therapy common distinct types generated generation internally	patients controls healthy training practice trained bipolar disorder subjects treatment patients therapy people transition person
Cluster_09	memory retrieval encoding	memory encoding retrieval	memory encoding retrieval	memory encoding retrieval

	object objects category face faces recognition motion visual perception emotional faces facial	events memory autobiographical stimulus visual repetition emotional negative emotion faces face facial	memory retrieval episodic context scenes scene object objects visual emotional faces facial	memory retrieval episodic faces emotional fearful object objects recognition items item memory
Cluster_10	emotional faces facial emotional negative emotion inhibition response inhibitory reward decision risk social participants empathy	reward anticipation monetary fear threat conditioning emotional negative emotion inhibition response control placebo dopamine healthy	reward monetary anticipation anxiety threat disorders emotion regulation emotional fear conditioning cs response responses hemodynamic	reward monetary anticipation emotion regulation emotional emotional processing neutral fear conditioning cs inhibition response control

Table 2

The topic mapping of subdivisions of the right thalamus from the dual view of NMF. Highlighting shows topics that are significant at $P < 0.01$ (uncorrected), and topics in red are significant at $P < 0.05$ (Bonferroni corrected).

right	topic 50	topic 100	topic 200	topic 400
Cluster_01	reward decision risk social participants empathy participants individual mechanisms task performance cognitive learning feedback error	decision choice decisions social empathy participants subject individual measures reasoning judgments moral correlated positive negative	decision choice decisions signals spatial subject reward monetary anticipation task performance activation social participants interactions	delayed discounting delay outcome probability expected sources source localization participants relative examined relationship relationships examine
Cluster_02	memory working wm patients controls reho language hemispheric lateralization children age adults number numerical arithmetic	reho risk regional sleep number numerical memory working task adhd asd autism time sustained duration	memory working verbal time duration timing task load difficulty reho regional rs wm memory working	time duration series reho regional resting sustained delay transient patients controls age selective selectivity haptic
Cluster_03	motor movement movements stimulation somatosensory tms conflict response trials pain painful stimulation learning feedback error	motor movement movements stimulation somatosensory tms sequences timing bimanual pain painful stimulation trials response trial	motor sensorimotor primary movements motor movement tms stimulation rtms pain painful stimulation stimulation somatosensory tactile	motor planning execution motor finger movements movement movements motor model parameters partial motor sensory areas
Cluster_04	connectivity functional network stimulus responses response participants individual mechanisms activation subjects condition women sex allele	connectivity functional network network default dmn areas activated network attention attentional visual imaging functional magnetic	network default dmn activations areas activation suppression life responses imaging functional magnetic mechanisms underlying understood	network default mode dmn network default pictures picture neutral information processing process autonomic rate arousal
Cluster_05	action actions motor eeg frequency ms visual auditory sensory sentences comprehension language semantic words word	action actions observation noise echo imaging visual auditory sensory eeg source ms functional areas lateral	auditory visual sensory sentences language comprehension mirror imitation system speech auditory sounds experience subjective physical	speech auditory production imitation human pifg events event occurring auditory sound sounds language languages linguistic
Cluster_06	object objects category learning feedback error pattern patterns ica action actions motor motion visual perception	semantic words word learning training performance reading words language action actions observation functional areas lateral	learning sequence implicit functional lateral organization modulation modulated influence spatial space navigation action actions observation	learning learned reinforcement semantic word words human features coding virtual navigation spatial abstract conceptual concrete
Cluster_07	emotional faces facial volume gray structural patients depression disorder emotional negative emotion training treatment practice	faces face facial volume gray structural depression mdd patients emotional negative emotion treatment baseline therapy	emotional faces facial volume gray structural anxiety threat disorders depression mdd patients treatment baseline therapy	volume gray structural faces emotional fearful facial expressions emotional risk risky taking disorders clinical symptoms
Cluster_08	patients schizophrenia controls patients depression disorder statistical functional subject training treatment practice patients disease ad	patients schizophrenia controls stress ptsd exposure pet flow tomography learning training performance depression mdd patients	schizophrenia patients controls he ocd patients depression mdd patients modulation modulated influence analyses roi cluster	schizophrenia patients controls depression mdd depressive treatment patients therapy controls compared healthy previous shown previously
Cluster_09	memory retrieval encoding object objects category	memory encoding retrieval stimulus visual repetition	memory encoding retrieval memory recognition recollection	memory encoding retrieval memory retrieval episodic

right	topic 50	topic 100	topic 200	topic 400
	motion visual perception	events memory autobiographical	memory retrieval episodic	recollection source familiarity
	eye spatial gaze	object objects visual	object objects visual	object objects recognition
	attention attentional visual	spatial space location	cognitive executive function	cognitive performance neuropsychological
Cluster_10	inhibition response inhibitory	reward anticipation monetary	emotion regulation emotional	reward monetary anticipation
	reward decision risk	feedback error errors	reward monetary anticipation	response responses increased
	food taste weight	inhibition response control	inhibition response inhibitory	mechanisms underlying understanding
	alcohol drug smokers	responses response imaging	mechanisms behavioral underlying	stress cortisol reactivity
	emotional negative emotion	experience subjective bias	evidence provide behavioral	ability identification individual

Table 3

Resting state functional connectivity of each thalamic cluster.

	Voxel size	MNI coordinate			Peak voxel	side	Identified brain regions
		x	y	z			
Cluster_01	2997	8	−4	8	28.4	R	thalamus (VA, AN)
	596	0	28	14	8.86	L	Anterior cingulate
	38	0	−16	−14	7.8	L	Red nucleus
Cluster_02	2122	−8	−10	14	37.83	L	thalamus(AN, VL)
	460	2	28	50	9.72	R	Surperior frontal gyrus
	294	−8	−80	−26	8.47	L	Declive
	279	32	−60	−28	7.79	R	Declive
	99	52	10	50	6.6	R	Middle frontal gyrus
	62	−48	14	38	6.86	L	Middle frontal gyrus
	34	−40	32	34	6.07	L	Middle frontal gyrus
	30	44	2	60	6.55	R	Middle frontal gyrus
	23	−22	−82	−24	6.35	L	Declive
	23	−44	50	4	5.75	L	Inferior Frontal gyrus
Cluster_03	2180	14	−14	8	43.95	R	thalamus(VL)
	2042	−14	−16	8	39.71	L	thalamus(VL, VPL)
	843	8	10	46	8.6	R	cingulate gyrus
	73	6	−70	16	6.7	R	Precuneus
	28	−16	−62	16	7.06	L	culmen
	25	62	4	20	6.46	R	precentral gyrus
	21	−8	−64	−16	6.04	L	culmen
	20	−8	−2	76	6.65	L	SMA
Cluster_04	753	4	−20	8	39.21	R	thalamus(MD)
	87	2	−22	50	6.38	R	medial frontal gyrus
	73	4	38	14	7.64	R	Anterior cingulate
	48	−2	0	54	6.88	L	Medial frontal gyrus
	43	−6	−74	36	7.08	L	Precuneus
	32	4	25	32	6.2	R	Anterior cingulate
	30	−2	4	46	6.27	L	middle cingulate
	23	−10	−66	22	6.1	L	Precuneus
Cluster_05	984	−52	−8	28	8.49	L	precentral gyrus
	687	14	−22	−2	37.47	R	thalamus(VP)
	629	24	−26	74	8.61	R	precentral gyrus

	Voxel size	MNI coordinate			Peak voxel	side	Identified brain regions
		x	y	z			
	627	2	−22	58	9.54	R	SMA
	356	62	−8	22	8.5	R	postcentral gyrus
	260	50	−30	22	7.63	R	Inferior Parietal gyrus
	85	−28	−18	−2	6.91	L	Lentiform Nucleus
	47	−58	−22	16	6.46	L	postcentral gyrus
	46	34	−14	−4	8.04	R	claustrum
	40	−62	−32	20	6.66	L	Superior temporal gyrus
	31	38	0	−4	6.66	R	Insula
	23	64	−22	20	6.46	R	postcentral gyrus
	22	−34	−22	50	6.18	L	postcentral gyrus
Cluster_06	3414	−26	−86	22	10.75	L	Cuneus
	402	20	−30	8	48.53	R	thalamus (pulvinar)
	340	−20	−28	8	49.17	L	thalamus (pulvinar)
	85	−22	−66	−10	7.72	L	Lingual gyrus
	73	22	−70	−10	7.63	R	Lingual gyrus
	73	−32	−70	−16	7.23	L	Fusiform gyrus
	34	−4	−58	68	7.22	L	precuneus
	30	−28	−58	58	6.35	L	Superior Parietal lobele
	20	28	−62	−4	6.3	R	Lingual gyrus
	20	10	−62	14	6.59	R	Posterior cingulate
Cluster_07	1179	−8	−32	4	49.28	L	thalamus (pulvinar)
	26	6	−50	10	6.5	R	Posterior cingulate
	20	−18	−22	4	8.37	L	thalamus(VPL)
	20	−48	−62	22	6.09	L	Middle temporal gyrus
Cluster_08	1288	−8	−56	70	8.71	L	postcentral gyrus
	1104	14	−22	14	44.84	R	thalamus(LD)
	201	20	−56	22	9.41	R	cingulate gyrus
	85	−12	−56	20	6.97	L	precuneus
	84	44	−76	32	7.41	R	angular gyrus
	70	−28	−44	−8	6.98	L	Parahippocampal gyrus
	60	34	−14	10	7.5	R	Insula
	54	−12	−32	44	7.13	L	cingulate gyrus
	52	38	−38	46	7.83	R	Inferior parietal lobule

	Voxel size	MNI coordinate			Peak voxel	side	Identified brain regions
		x	y	z			
	49	30	−40	−10	8.32	R	Parahippocampal gyrus
	46	50	−26	40	6.56	R	Postcentral gyrus
	42	52	−62	−6	6.71	R	Inferior temporal gyrus
	41	40	−44	58	6.49	R	Inferior parietal lobule
	38	44	−38	64	7.25	R	Postcentral gyrus
	32	52	−26	52	6.71	R	Postcentral gyrus
	30	4	−40	50	6.28	R	precuneus
	26	−52	−70	2	5.79	L	middle occipital gyrus
	26	34	−78	40	6.6	R	precuneus
	24	−44	−38	64	6.36	L	postcentral gyrus
	22	−34	−20	14	8.01	L	Insula
Cluster_09	13459	20	−28	−4	44.4	R	thalamus(LGN)
	707	8	44	−8	8.81	R	Middle orbital gyrus
	125	−44	8	−14	9.11	L	Superior temporal gyrus
	92	4	−34	38	6.96	R	Middle cingulate gyrus
	87	2	−56	−28	7.34	R	Culmen
	81	2	−10	10	9.59	R	thalamus(MD)
	68	28	−80	32	3.89	R	Cuneus
	49	58	2	−18	6.71	R	middle temporal gyrus
	40	−26	−80	32	6.2	L	Cuneus
	29	18	−40	−46	7.25	R	Cerebellar Tonsil
	28	−20	−36	−44	6.61	L	Cerebellar Tonsil
	26	32	32	−14	6.51	R	Inferior Frontal gyrus
	25	−58	−4	−12	5.86	L	middle temporal gyrus
	22	−56	−10	−10	7	L	middle temporal gyrus
	21	−28	30	−16	6.51	L	Inferior Frontal gyrus
Cluster_10	1915	8	−16	−4	32.45	R	thalamus(Red nucleus)
	643	−34	−2	−10	10.54	L	Claustrum
	250	40	14	−10	8.63	R	insula
	144	10	28	32	7.69	R	Anterior cingulate cortex
	71	32	−8	−10	7.16	R	Amygdala
	49	−20	−4	−38	7.54	L	Uncus

Voxel size	MNI coordinate			Peak voxel	side	Identified brain regions
	x	y	z			
49	24	−38	−2	8.36	R	Hippocampus
32	34	−10	−32	6.46	R	Fusiform gyrus