Presurgical Brain Mapping of the Language Network in Patients With Brain Tumors Using Resting-State fMRI: Comparison With Task fMRI

Haris I. Sair,[1]* Noushin Yahyavi-Firouz-Abadi,[1] Vince D. Calhoun,[2]
Raag D. Airan,[1] Shruti Agarwal,[1] Jarunee Intrapiromkul,[1] Ann S. Choe,[3]
Sachin K. Gujar,[1] Brian Caffo,[4] Martin A. Lindquist,[4] and Jay J. Pillai[1]

[1]Division of Neuroradiology, the Russell H. Morgan Department of Radiology and
Radiological Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
[2]The Mind Research Network, Departments of Electrical and Computer Engineering,
University of New Mexico, Albuquerque, New Mexico
[3]F. M. Kirby Research Center for Functional Brain Imaging, Kennedy Krieger Institute,
Baltimore, Maryland
[4]Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland

◆ ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

Abstract: *Purpose:* To compare language networks derived from resting-state fMRI (rs-fMRI) with task-fMRI in patients with brain tumors and investigate variables that affect rs-fMRI vs task-fMRI concordance. *Materials and Methods:* Independent component analysis (ICA) of rs-fMRI was performed with 20, 30, 40, and 50 target components (ICA20 to ICA50) and language networks identified for patients presenting for presurgical fMRI mapping between 1/1/2009 and 7/1/2015. 49 patients were analyzed fulfilling criteria for presence of brain tumors, no prior brain surgery, and adequate task-fMRI performance. Rs-vs-task-fMRI concordance was measured using Dice coefficients across varying fMRI thresholds before and after noise removal. Multi-thresholded Dice coefficient volume under the surface (DiceVUS) and maximum Dice coefficient (MaxDice) were calculated. One-way Analysis of Variance (ANOVA) was performed to determine significance of DiceVUS and MaxDice between the four ICA order groups. Age, Sex, Handedness, Tumor Side, Tumor Size, WHO Grade, number of scrubbed volumes, image intensity root mean square (iRMS), and mean framewise displacement (FD) were used as predictors for VUS in a linear regression. *Results:* Artificial elevation of rs-fMRI vs task-fMRI concordance is seen at low thresholds due to noise. Noise-removed group-mean DiceVUS and MaxDice improved as ICA order increased, however ANOVA demonstrated no statistically significant difference between the four groups. Linear regression demonstrated an association between iRMS and DiceVUS for ICA30-50, and iRMS and MaxDice for ICA50. *Conclusion:* Overall there is moderate group level rs-vs-task fMRI language network concordance, however substantial subject-level variability exists; iRMS may be used to determine reliability of rs-fMRI derived language networks. *Hum Brain Mapp 37:913–923, 2016.* © 2015 Wiley Periodicals, Inc.

Key words: resting-state fMRI; task-fMRI; language network; presurgical brain mapping; brain tumor
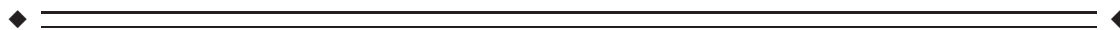
◆ ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

INTRODUCTION

Mapping of eloquent brain areas using functional magnetic resonance imaging (fMRI) is widely used for preoperative planning. There are varying accounts on the accuracy of fMRI language mapping [Bizzi et al., 2008; Giussani et al., 2010; Kuchcinski et al., 2015; Rutten et al., 2002]. Nevertheless, the use of fMRI can result in reduced surgical time, increased extent of resection, and reduced craniotomy size [Petrella et al., 2006], as well as accurate prediction of perioperative deficits [Bailey et al., 2015; Pillai, 2010]. Furthermore, fMRI demonstrates high accuracy in determining hemispheric language dominance, aiding the neurosurgeon in assessing operative risk [Dym et al., 2011]. Indeed, in many institutions, fMRI may be considered standard of care in the workup of patients for operative planning when the trajectory or resection involves critical brain regions [Gabriel et al., 2014].

There are, however, several limitations of fMRI as it is currently implemented. Brain activation is determined by correlation of changes in blood oxygenation level dependent (BOLD) signal with performance of a task that is modulated across fMRI acquisition [Ogawa et al., 1993]. As direct neuronal activation is not being measured, a hemodynamic response function (HRF) is used to model task-related changes in BOLD signal. The task paradigm must be constructed carefully to ensure that the design accurately activates brain regions for a specific function, while minimizing inclusion of confounding neurobehavioral functions. Generally, for language mapping, a combination of multiple tasks is needed to ensure activation of the global language network. Specialized stimulus presentation equipment and software are also needed for implementation of the paradigms, and highly trained personnel are needed to assess the patient's cognitive status prior to fMRI scanning, select appropriate paradigms and evaluate task performance. All the considerations aforementioned contribute to added cost in the patient's medical care. In addition, ensuring adequate task compliance may be challenging in neurologically debilitated patients or in the pediatric population.

Resting-state fMRI (rs-fMRI) has the potential to overcome some of these limitations. In rs-fMRI, no explicit task is given, and spontaneous fluctuations in BOLD signal across brain regions are measured [Fox et al., 2005]. This technique is attractive as it overcomes two common shortcomings of task-fMRI: the necessity to model the HRF to estimate task-related brain activity and the need for patient compliance and capability in performing a task.

In rs-fMRI, spatially distinct but temporally synchronous regions of signal changes are seen representing intrinsic brain networks, with the motor network universally and the language network variably being demonstrated in group studies [Allen et al., 2011; Damoiseaux et al., 2006; Kalcher et al., 2012]. Several studies have examined the potential use of rs-fMRI in presurgical brain mapping. There is a high overlap between rs-fMRI and task-fMRI when comparing the motor network [Kokkonen et al., 2009; Shimony et al., 2009; Zhang et al., 2009], with rs-fMRI showing high concordance with cortical stimulation mapping [Zhang et al., 2009]. Reproducibility of rs-fMRI-derived motor maps is also comparable to that of task-fMRI [Mannfolk et al., 2011]. There is moderate overlap of the language network in healthy subjects between rs-fMRI and task-fMRI [Tie et al., 2014] with high reproducibility [Kollndorfer et al., 2013]. Mitchell et al. [2013] demonstrated in a limited number of subjects that an artificial neural network algorithm can reliably identify motor and language networks even in cases of grossly altered anatomy due to lesion mass effect.

In this study, we investigated the concordance of rs-fMRI-derived language networks with task-fMRI language activation in a cohort of patients with brain tumors presenting for presurgical mapping. We hypothesized that there is variability in the concordance of rs-fMRI and task-fMRI across subjects and explored variables that may impact the reliability of rs-fMRI for presurgical mapping.

MATERIALS AND METHODS

Participants

For this IRB-approved retrospective study, the Johns Hopkins Hospital radiology information system (RIS) was interrogated for any patient who underwent fMRI for presurgical brain mapping between 1/1/2009 and 7/1/2015. Seventy-nine patients had rs-fMRI in addition to task-fMRI during the same imaging session. Twenty-one of these patients had prior history of brain surgery (including biopsy) and were excluded. One patient demonstrated imaging findings consistent with an intracranial arteriovenous malformation (AVM) and was excluded. Language fMRI for presurgical brain mapping is tailored to each individual patient in several ways. First, a task must be selected that is within the performance capabilities of the patient (e.g., a language task based on visual cues will not be used on a patient who is not able to adequately see the stimulus). Second, the location of the lesion influences the type of task chosen as the goal is to identify language related brain areas along the expected surgical trajectory and in the immediate vicinity of the lesion. For a lesion near the posterior temporal lobe for example, a receptive language task that robustly activates Wernicke's area is necessary. For this study, we selected patients who had performed three tasks that are most commonly utilized in combination at our institution for assessment of global language regions. The three tasks were rhyming (Rhym), sentence completion (SC), and silent word generation (SWG). Five patients who did not complete all three tasks were excluded. Of the remaining fifty-two patients, three patients were excluded due to lack of reliable language-related activation on task-fMRI data, which was assessed qualitatively by a subspecialty board-certified

neuroradiologist with more than 5 years of experience in presurgical brain mapping using fMRI. Forty-nine patients were thus included for final analysis (age range 18–69/mean 39.82 years; 31 males and 18 females).

Handedness

A laterality index (LI) for patient handedness was calculated using the Edinburgh Handedness Inventory questionnaire [Oldfield, 1971]. LI was calculated in a similar manner to calculation of LI for hemispheric language dominance in fMRI:

$$\frac{\text{Number of responses RIGHT - Number of responses LEFT}}{\text{Number of responses RIGHT + Number of responses LEFT}}$$

Lesion Characterization

For each patient, lesion location (left vs right hemisphere), volume (in cubic millimeters), and pathology (when available) were documented. For brain tumors, the World Health Organization (WHO) histologic grade was recorded. Lesion volume for each subject was calculated by manual region-of-interest (ROI) drawing on T2-Fluid-attenuated inversion recovery (FLAIR) images (parameters below) in the Medical Image Processing, Analysis, and Visualization (MIPAV) application (National Institutes of Health, Bethesda, Maryland). For high-grade brain tumors, MRI cannot assess the true extent of neoplasm, and there are fractional contributions of edema and infiltrative neoplasm in areas of T2 signal hyperintensity. As our primary interest was spatial extent of abnormality as a variable, we included all areas of lesion-related T2 signal hyperintensity as the final lesion volume.

Imaging

MR images were obtained using a 3.0 T Siemens Trio Tim system (Siemens Medical Solutions, Erlangen, Germany) with a 12-channel head matrix coil. Structural images included a three-dimensional (3D) T1 sequence (TR = 2300 ms, TI = 900 ms, TE = 3.5 ms, flip angle = 9°, field of view = 24 cm, acquisition matrix = 256 × 256 × 176, slice thickness = 1 mm) and a two-dimensional (2D) T2-FLAIR sequence (TR = 9310 ms, TI = 2500 ms, TE = 116 ms, flip angle = 141°, field of view = 24 cm, acquisition matrix = 320 × 240 × 50, slice thickness = 3 mm). Functional T2*-weighted BOLD images for both task-fMRI and rs-fMRI were acquired using 2D gradient-echo echo-planar imaging (TR = 2000 ms, TE = 30 ms, flip angle = 90°, field of view = 24 cm, acquisition matrix = 64 × 64 × 33, slice thickness = 4 mm, slice gap = 1 mm, interleaved acquisition). For rs-fMRI, 180 volumes were acquired (6 min), with three instructions: (1) try to keep still, (2) keep your eyes closed, and (3) do not fall asleep.

Task-fMRI Paradigms

For each patient, instructions and practice sessions were provided outside the scanner on each task to be performed. During acquisition, real-time fMRI maps for each task were monitored by the neuroradiologist to assess for global data quality; any task performance that was deemed suboptimal due to motion-related or other artifact was repeated per our protocol. Language fMRI tasks were implemented using stimulus presentation software provided by Prism Clinical Imaging (Elm Grove, Wisconsin).

All paradigms used a block design of either 30 s (Rhym) or 20 s (SC, SWG) alternating task and control blocks for total acquisition duration of 3 or 4 min, respectively [Pillai and Zaca, 2011; Zacà et al., 2012, 2013].

Image Processing

The fMRI imaging data was processed using Statistical Parametric Mapping (SPM) version 8 (Wellcome Department of Imaging Neuroscience, University College London, UK) and custom MATLAB (Mathworks, Natick, Massachusetts) scripts.

Task-fMRI underwent slice timing correction (STC) followed by motion correction (MC), normalization to a Montreal Neurological Institute (MNI-152) template, and spatial smoothing using a 6 mm full width at half maximum (FWHM) Gaussian kernel.

Rs-fMRI underwent STC followed by MC. The ArtRepair toolbox [Mazaika et al., 2009] was then used to detect volumes with large shifts in global average signal intensity related to scan-to-scan motion; both the outlier volumes and additional volumes recommended for deweighting were tagged for subsequent removal from analysis (i.e., for "scrubbing"). Rs-fMRI was linearly detrended. Following coregistration of rs-fMRI and T1-weighted images, physiological nuisance regression of rs-fMRI was performed utilizing the CompCor method [Behzadi et al., 2007] using signal extracted from eroded white matter and cerebrospinal fluid masks. After bandpass filtering from 0.01 to 0.1 Hz, smoothing was performed with a 6 mm FWHM Gaussian kernel. Finally, images tagged by ArtRepair were removed ("scrubbed") from the rs-fMRI volumes.

For each subject, the variance of signal intensity across time was measured by calculating the average root mean square fluctuation in signal intensity of all voxels within subject-specific brain masks (image intensity Root Mean Squared, or iRMS, denoted as percent signal change). In addition, mean volume-to-volume framewise displacement (FD) was calculated per Power et al. [2012].

All task-fMRI and rs-fMRI images were visually inspected throughout the different steps of processing for quality control.
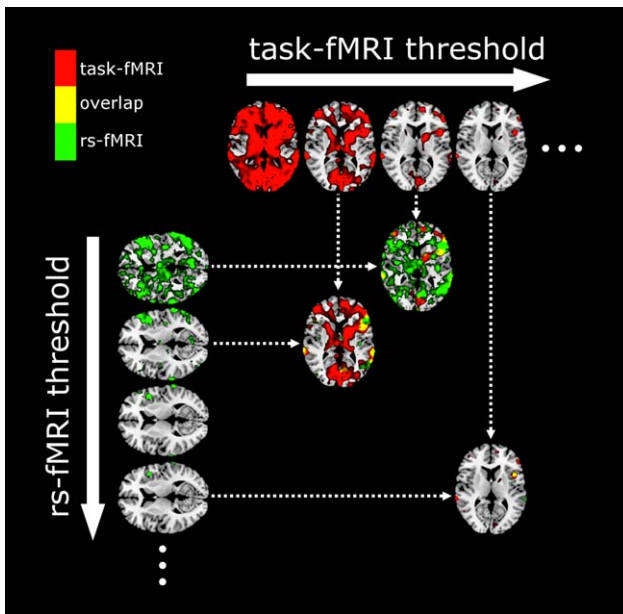
Figure 1.

Dice coefficient calculation across multiple thresholds. For each subject, normalized task-fMRI and rs-fMRI were thresholded from 0 to 10 at 0.1 increments, and Dice coefficients calculated across all thresholds to generate a 100 × 100 matrix of Dice coefficients. The figure demonstrates brain maps corresponding to four random levels of thresholding on the *x* (task-fMRI) and *y* (rs-fMRI) axes, and three selected examples of spatial overlap maps as the target of the arrows to demonstrate the contribution of noise in Dice calculation at low thresholds. The ellipses denote the presence of additional thresholds that are not shown. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

IMAGE ANALYSIS

Task-fMRI

General linear model (GLM) analysis was used for task-fMRI using SPM8 using the canonical HRF convolved with the boxcar function for each task with the following parameters: no model derivatives; 128 s high-pass filter; no global intensity normalization; autoregressive model (AR1) to account for temporal autocorrelations; and no confound matrix. A design matrix was constructed for each patient with the contrast set to detect activation across all three tasks compared to rest. The SPM T-contrast maps were generated without clustering or multiple comparison correction; the former was not performed due to the necessity of multilevel thresholding (below), and the latter was not performed as it would result in overly stringent thresholding for clinical language activation maps.

Rs-fMRI

Rs-fMRI was analyzed using the group independent component analysis (ICA) of fMRI Toolbox (GIFT, Medical Image Analysis Lab, http://mialab.mrn.org/software/gift). ICA was performed separately for each subject using the InfoMax algorithm with ICASSO set at 5 repeats to determine reliability of the ICA maps [Himberg et al., 2004]. ICA maps were generated for 20, 30, 40, and 50 components, designated here as ICA20, ICA30, ICA40, and ICA50, respectively. For each ICA order, to minimize errors in identification of the language network, spatial sorting was performed using the multiple regression method implemented in GIFT and the task-fMRI T-maps used as the reference template. Because the level of task-fMRI T-map thresholding may affect the order of the best matched rs-fMRI ICA component (e.g., at low thresholds, more noise elements are present in the task-fMRI which could result in selection of noise rs-fMRI ICA components as the best match, and at high thresholds, limited active voxels are present to determine the best rs-fMRI match), we utilized 4 different levels of task-fMRI thresholding after signal normalization (below) as the target maps; the 4 levels represented thresholds at 20%, 40%, 60%, and 80% of the maximum fMRI signal value across the whole brain. The top three rs-fMRI ICA components for each thresholded task-fMRI target were then visually inspected to select the one component for each ICA order that demonstrated the highest spatial overlap with the task-fMRI maps; this represented the best candidate for the rs-fMRI language network. Because the level of noise in the task-fMRI activation maps may differ across subjects for a particular threshold level, we qualitatively assessed the target task-fMRI activation maps to weight the rs-fMRI network selection to the task-fMRI threshold demonstrating the least amount of noise. As an example, consider a condition where the multiple regression algorithm listed rs-fMRI component 5 as the best match for task-fMRI thresholded at 20% and 40% and component 11 as the best match for task-fMRI thresholded at 60% and 80%. If qualitative review of the task-fMRI shows noise at the low levels of thresholds (20% and 40%) with no clear localization of language-specific activation, however, clear language-specific activation is seen at 60% and 80% thresholding, then component 11 rather than component 5 in this case is chosen as the best candidate for the rs-fMRI language network for subsequent analysis.

For each subject, we conservatively opted to force selection of a "best candidate" language network even when a clear language network was not readily identified. Discarding data where the rs-fMRI language network was not reliably identified would limit subsequent analysis to datasets with good rs-task correlations, and would then artificially inflate the value of rs-task correlations.

FMRI Comparison

The choice of thresholding to determine regions of fMRI activation for a single subject is somewhat arbitrary, with low reproducibility at a given threshold even within a subject across scans [Voyvodic, 2012]. With a low threshold,
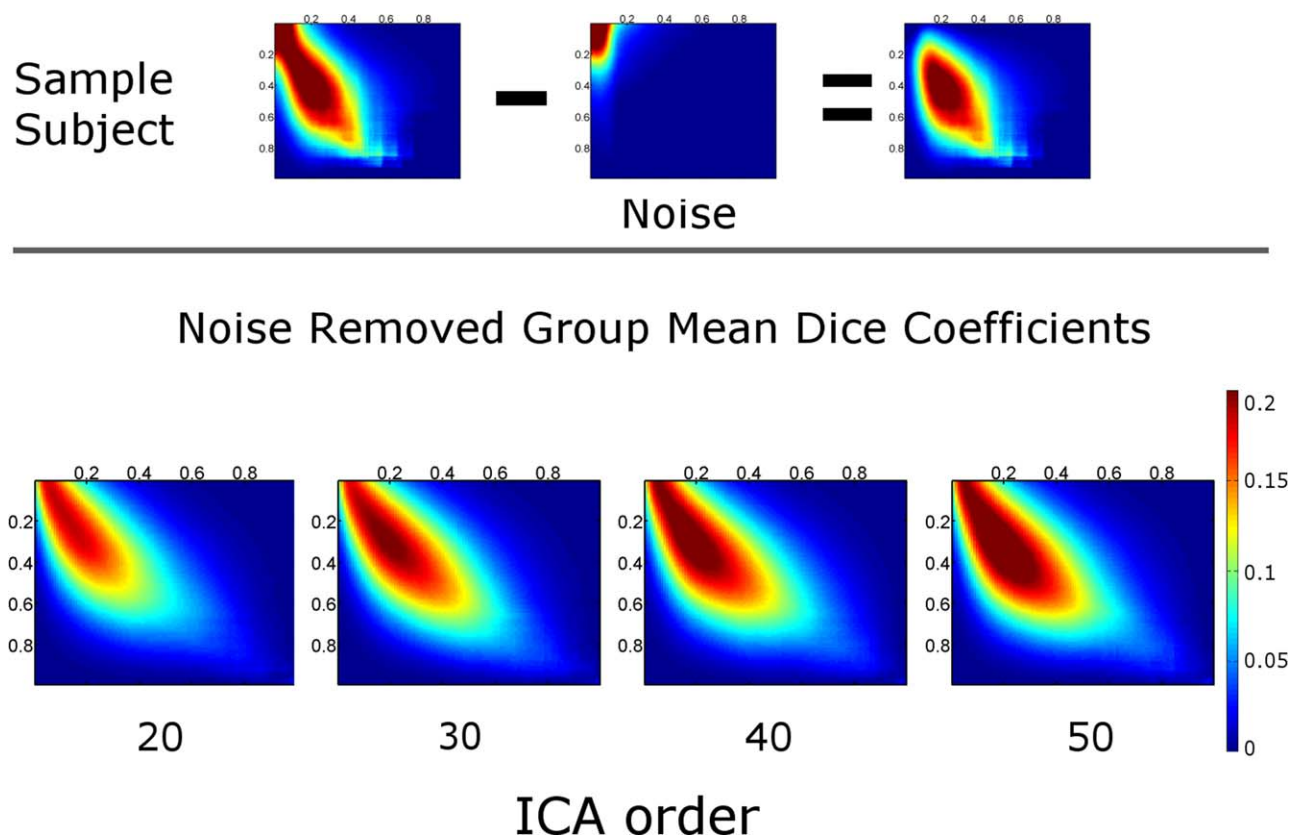
Figure 2.

Dice coefficient matrix. On top, a Dice coefficient map across multiple thresholds is shown for a single subject. From the raw Dice coefficient map (left image), a noise map (middle image) is subtracted to generate a noise-removed Dice coefficient map (right image). Maps were averaged across all subjects for each ICA order and noise-removed Group mean Dice coefficient maps are shown on the bottom. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

rs-task concordance will be artificially elevated due to inclusion of noise elements. To overcome this issue, we normalized the task-fMRI T-maps and the rs-fMRI ICA maps as a fraction of maximum signal intensity across all voxels to a value of 0–10 following elimination of negative values, and thresholded each map at 0.1 increments. For each of the 100 rs-fMRI and 100 task-fMRI thresholds, Dice coefficients D between rs-fMRI and task-fMRI were calculated to generate a 100×100 matrix of task-fMRI (t) vs rs-fMRI (r) overlap (Fig. 1):

$$D = \frac{2|t \cap r|}{|t| + |r|}$$

At low thresholds, high Dice coefficients result from overlap of random noise. We calculated the volume under the surface (VUS) and the maximum Dice coefficient (Max-Dice) for each Dice coefficient map before and after noise removal. For calculation of noise, we selected an ICA component representing anterior ventricular signal for each sub-ject, and performed intensity normalization and multithresholding after removal of negative values. Dice coefficients between the noise and task maps were calcu-lated. The resultant noise-vs-task Dice map was subtracted from the rs-vs-task Dice map, and finally noise-corrected rs-vs-task Dice map VUS (DiceVUS) and MaxDice values were calculated (Fig. 2, top). One-way analysis of variance (ANOVA) was performed to determine significant differen-ces in DiceVUS and maxDice across the four ICA orders.

Predictor of DiceVUS

For each ICA order, a linear regression was performed using R [R Core Team, 2015] to determine if there were any predictors of DiceVUS and maxDice. The predictors were: age, sex, LI, tumor side, tumor size, WHO grade, number of scrubbed volumes, image intensity root mean square percent variation (iRMS%), and mean framewise displacement (FD).

TABLE I. Summary of patient data

		Tumor type
Age	39.8/18-69	Glioblastoma (9)
Sex	31M/18F	Oligodendroglioma (10)
Handedness LI	0.61/-0.91-1.00	Anaplastic astrocytoma (7)
		Oligoastrocytoma (5)
		Diffuse glioma (5)
		Infiltrating astrocytoma with early anaplastic transformation (4)
		Anaplastic oligoastrocytoma (1)
		Gangliocytoma (1)
		Glioblastoma with oligodendroglial component (1)
		Diffuse large B-cell lymphoma (1)
WHO grade		
I	1	
II	20	No pathology:
III	12	Likely low-grade glioma on imaging (5)
IV	10	

RESULTS

Handedness

LI ranged from −0.91 (strongly left handed) to 1 (exclusively right handed), with mean of 0.61 (right hand preference). Using a LI cutoff of ±0.2 from zero (completely ambidextrous), there were 41 right-handed patients, 6 left-handed patients, and 2 ambidextrous patients.

Lesions

Histopathology and WHO grade of lesions included in this study are listed in Table I. Most lesions were primarily located in the left cerebral hemisphere (38 lesions). Eleven lesions were primarily in the right cerebral hemisphere. Tumor size ranged from 1.02×10^3 mm^3 to 1.59×10^5 mm^3 (mean 4.12×10^4 mm^3).

Rs-fMRI

iRMS% ranged from 1.41 to 4.04 (mean 2.22). ArtRepair identified from 1 to 101 outlier volumes (mean 8.8 volumes) which were tagged for scrubbing. Mean FD ranged from 0.05 to 0.42 mm (mean 0.16).

FMRI Comparison

Group mean Dice maps for each ICA order are shown in Figure 2, and sample images from three subjects with varying levels of Dice coefficients are shown in Figure 3.
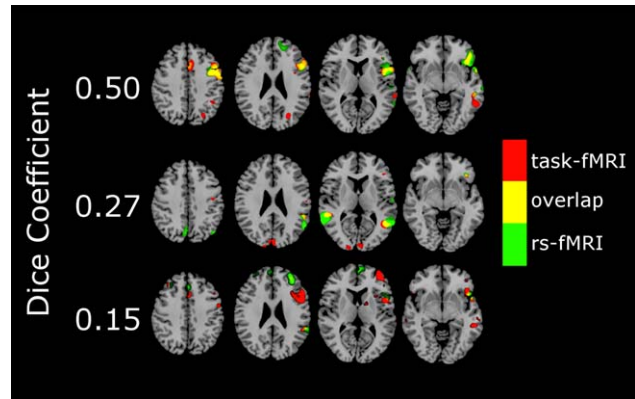
Figure 3.

Sample subjects at varying Dice coefficients. Each row denotes a different subject with fMRI overlays onto standard brain template for comparison. Task-fMRI and rs-fMRI were thresholded at the peak Dice coefficient value for each of these subjects. Task-fMRI is shown in red, the rs-fMRI in green, and the overlapping areas in yellow. For the subject in the top-most row, there is an excellent concordance of task-fMRI and rs-fMRI. For the subject in the bottom-most row, there is a low concordance of task-fMRI and rs-fMRI, reflected in the low Dice coefficient. The subject in the middle row demonstrates an intermediate level of task-fMRI and rs-fMRI concordance. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Group mean DiceVUS overall increased with ICA order (Fig. 4). Mean DiceVUS was 425 for ICA20, 513 for ICA30, 546 for ICA40, and 566 for ICA50. One-way ANOVA demonstrated no significant differences in DiceVUS among the four ICA groups (Table II). Mean MaxDice was 0.28 for ICA20, 0.31 for ICA30, 0.33 for ICA40, and 0.33 for ICA50,
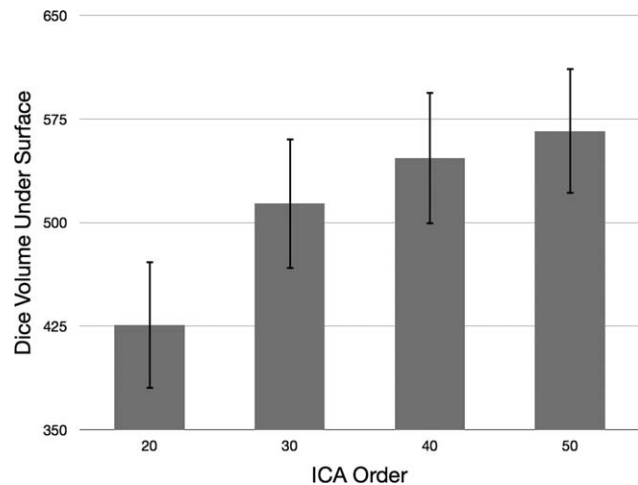
Figure 4.

Mean DiceVUS across ICA orders shown with standard error bars. Mean DiceVUS overall increases as ICA order increases; however, this effect does not reach statistical significance.

TABLE II. ANOVA and Tukey's Honest Significant Difference (HSD) between the four ICA orders for DiceVUS

DiceVUS

	Sum of Squares	Df	Mean Square	F	P
Between groups	567830	3	189277	1.76	0.156

	Mean difference	Lower Bound	Upper bound	p adj
ICA30 vs ICA20	88.02	−83.70	259.73	0.55
ICA40 vs ICA20	121.04	−50.67	292.76	0.26
ICA50 vs ICA20	140.48	−31.23	312.19	0.15
ICA40 vs ICA30	33.03	−138.68	204.74	0.96
ICA50 vs ICA30	52.46	−119.25	224.18	0.86
ICA50 vs ICA40	19.43	−152.28	191.15	0.99

No significant differences were found in DiceVUS across the four groups. Lower and upper bounds are at 95% confidence interval.

with no significant differences among the four ICA groups using one-way ANOVA (Table III and Fig. 5).

Predictor of DiceVUS and MaxDice

iRMS was predictive of DiceVUS for ICA30, ICA40, and ICA50 (Table IV). iRMS trended toward significance for ICA20. LI was predictive of DiceVUS for ICA30. iRMS was predictive of MaxDice for only ICA50.

DISCUSSION

As rs-fMRI continues to gain popularity as a novel method of examining brain function, considerable effort is being made to apply this technique in a clinically relevant fashion, from disease diagnosis to management [Castellanos et al., 2013; Lee et al., 2013]. Previous studies explored the feasibility of rs-fMRI in presurgical brain mapping [Qiu et al., 2014; Shimony et al., 2009; Tie et al., 2014]. Our study directly compares task-fMRI to rs-fMRI in brain tumor patients who presented for presurgical language mapping, and to date (and our knowledge) represents the largest cohort of this kind.

Although there is an excellent language network concordance between rs-fMRI and task-fMRI in some subjects, we demonstrate significant variability across subjects. We found that the percent root mean square of image intensity (iRMS) was predictive of concordance in three out of four ICA orders (ICA30-ICA50) and trended toward significance in the remaining ICA order (ICA20). FD, however, predicted neither DiceVUS nor MaxDice. Variations in image intensity (primarily attributed to motion) rather than absolute measurements of movement, therefore, appear to be more sensitive to data quality, and may be used as a general marker of rs-fMRI reliability for this purpose. An alternative related method of calculating intensity-based variance is the DVARS ("D" = temporal derivative of time courses, "VARS" = RMS variance over voxels) method [Power et al., 2012], which measures the

TABLE III. ANOVA and Tukey's Honest Significant Difference (HSD) between the four ICA orders for MaxDice

MaxDice

	Sum of squares	Df	Mean square	F	p
Between groups	0.08	3	0.03	1.22	0.303

	Mean difference	Lower Bound	Upper bound	p adj
ICA30 vs ICA20	0.03	−0.05	0.10	0.80
ICA40 vs ICA20	0.05	−0.03	0.13	0.36
ICA50 vs ICA20	0.05	−0.03	0.13	0.36
ICA40 vs ICA30	0.02	−0.06	0.10	0.89
ICA50 vs ICA30	0.02	−0.06	0.10	0.89
ICA50 vs ICA40	−0.00002	−0.07	0.07	1.0

No significant differences were found in MaxDice across the four groups. Lower and upper bounds are at 95% confidence interval.
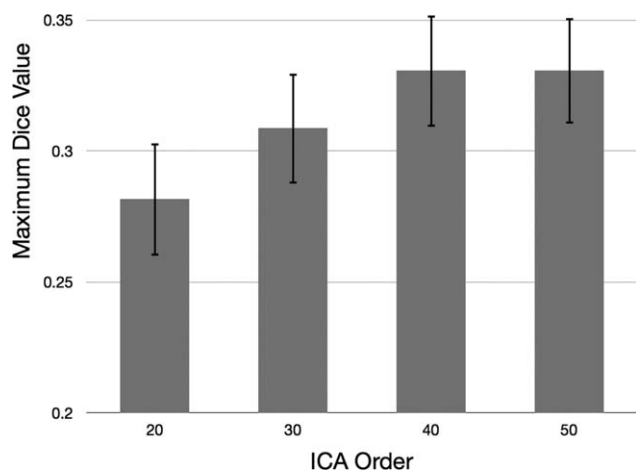
Figure 5.

Mean MaxDice coefficients across ICA orders shown with standard error. MaxDice increases from ICA20 to ICA40; however, this effect does not reach statistical significance.

rate of change of BOLD signal between two successive fMRI volumes. We opted to use the iRMS method as it is implemented in a well-utilized toolbox, and there is variability in the exact method of calculation of DVARS (e.g., the method of intensity normalization, or the step at which a brain mask is utilized, both of which can vary depending on the program used).

We demonstrate that the number of independent components chosen in this study has no significant effect on rs-task concordance. Although the group mean Dice coefficient increased slightly as ICA order increased, ANOVA demonstrated no significant differences across the four ICA groups at least in part due to high subject-level variability of DiceVUS. Estimation of the ideal number of independent components poses a unique challenge in fMRI due to the spatial and temporal dependence of BOLD signal, with decreased repeated-measure stability of

independent component estimates using higher orders [Li et al., 2007]. Under/overfitting can result in merging of different brain networks at low orders and fragmentation of canonical networks into subnetworks at high orders. Various methods of optimal ICA order estimation using information theory criteria (ITC) exist. Although a study comparing various methods for order estimation concluded that no perfect ITC can be determined for fMRI data [Hui et al., 2011], using subject-specific ICA order estimation may potentially improve the results and warrants future investigation.

We also demonstrate the effect of artificial elevation of Dice coefficients at low thresholds of fMRI due to noise, and implement a noise-removal method. The noise elements do not significantly change across ICA orders, and substantially contribute to Dice measurements. These artificially elevated values will overestimate the potential of rs-fMRI. Therefore, inclusion of noise removal steps and calculating concordance using various thresholds is recommended for future studies.

The choice of rs-fMRI processing technique may affect connectivity estimates. A core aspect of rs-fMRI processing involves attempting to diminish the effect of motion and physiological noise [Birn, 2012]. Motion has a significant effect on connectivity measurements [Van Dijk et al., 2012]. Scrubbing can be performed to censor high motion volumes [Power et al., 2012, 2014]. We opted to utilize a range that has been typically used in prior studies of this kind. Although we find that the number of outlier volumes has no effect on concordance, nevertheless, increasing scan length may improve rs-fMRI quality [Birn et al., 2013].

Seed-based analysis is an alternative to ICA analysis of rs-fMRI. We did not explore seed-based analysis in this initial evaluation of this dataset for various reasons. Universal seed locations across subjects cannot be used in this population due to the presence of brain lesions with mass effect. This may be partially ameliorated with nonlinear

TABLE IV. Variables associated with DiceVUS and MaxDice across the four ICA orders (p values) using linear regression

	Dice VUS				MaxDice			
	ICA order							
	20	30	40	50	20	30	40	50
Age	0.23	0.21	0.20	0.23	0.42	0.21	0.11	0.19
Sex	0.71	0.40	0.60	0.49	0.73	0.44	0.54	0.76
LI	0.16	**0.02***	0.07	0.10	0.90	0.43	0.49	0.56
Tumor Side	0.87	0.27	0.09	0.42	0.58	0.32	0.17	0.28
Tumor Size	0.75	0.60	0.52	0.58	0.89	0.82	0.71	0.69
WHO Grade	0.27	0.78	0.23	0.58	0.27	0.54	0.57	0.52
#scrubbed	0.80	0.27	0.24	0.15	0.61	0.14	0.24	0.06
iRMS	0.08	**0.04***	**0.02***	**0.006***	0.34	0.08	0.13	**0.02***
Mean FD	0.43	0.15	0.13	0.12	0.27	0.14	0.10	0.12

*$p < 0.05$.

normalization with lesion masking. While seed-based analyses have shown promise in outlining motor networks in both healthy controls [Kristo et al., 2014] and patients with brain lesions [Rosazza et al., 2014], the choice of seed placement is much more challenging for determination of language networks due to the greater degree of intersubject variability in language functional anatomy [Sanai et al., 2008], which precludes ROI placement based solely on *a priori* knowledge of classical language representation areas. The motor cortex is relatively easily identified even in patients with large brain shifts, and functional variability of the motor cortex does not extend beyond motor network subcomponents. On the other hand, the inferior frontal gyrus (IFG) includes various functional subunits [Xiang et al., 2010], and accurate determination of receptive language areas along the posterior temporal lobe and inferior parietal lobule is difficult based on anatomical landmarks. As minor differences in seed placement can significantly affect network metrics [Cole et al., 2010], this method remains challenging for accurate characterization of the language network, and for rs-fMRI to be used as a clinically viable tool for presurgical brain mapping, any element that introduces operator-dependent error needs to be minimized. A potentially useful study utilizing our dataset would be to use coordinates of productive and/or receptive language activation from each subject's task-fMRI and use those coordinates as seeds for generating the rs-fMRI language network. This would determine the maximum inherent information on the subject's language network that can be extracted using rs-fMRI. If this method yields superior rs-task concordance, it would mean that the ICA method needs to be optimized. On the other hand, if the concordance is similar to what is shown in this study, then it represents the best information that is available using these acquisition parameters, and modification/optimization of other variables such as scanner field strength or length of acquisition need to be explored.

Several limitations of task-fMRI analysis in this study need to be addressed. Although we combined three of the most commonly utilized tasks at our institution to maximize sensitivity and specificity for detecting language areas, various non-language-specific cortical activations can be commonly seen which may affect the overall rs-task concordance. Depending on the strength of activation of these regions, the rs-task concordance may be lowered. However, exclusion of these so-called non-primary language areas is also problematic due to the variable contribution of non-language-specific regions to language processing across subjects.

ICA has also been used to analyze task-fMRI data, and potentially could separate language-specific vs non-language-specific areas; however, in our experience we find that at the single-subject level, the β-maps of task-fMRI ICA analysis can also be significantly affected by areas outside of primary language networks. Some of these findings may be explained by the nonstationary (dynamic) nature of functional connectivity, with intrinsic networks engaging with other networks across time [Chang and Glover, 2010; Fedorenko and Thompson-Schill, 2014; Hutchison et al., 2013]. Future studies on spatially constrained ICA and its variants may be useful here [Lin et al., 2010]. As ICA analysis of task-fMRI is far less commonly utilized (if ever) in actual clinical practice, we chose to base our analysis on GLM maps.

Another important limitation in this study is the fact that rs-fMRI was performed after task-fMRI and functional connectivity can be influenced by prior tasks [de Chastelaine and Rugg, 2014; Wang et al., 2012]. This arrangement was necessary in this clinical cohort to ensure that the patients did not fatigue from the length of the scan that could compromise their performance on task-fMRI. In some patients who may have been unable to tolerate longer scan times, rs-fMRI acquisition may have been omitted, potentially introducing another source of bias.

While we specifically instructed the patients to stay awake during the rs-fMRI, there was no systematic way of ensuring that this instruction was followed. Although postscan questionnaires can be utilized, no such information was available for this retrospective study. This is a significant limitation as alterations in connectivity have been described in sleep, or even with varying levels of vigilance [Tagliazucchi and Laufs, 2014]. To date there is no systematic study specifically detailing the effect of sleep on the intrinsic language network. A single case report demonstrates receptive language task activation during sleep in a 6-year-old child that was concordant with areas of language activation while awake [Wilke et al., 2003]; this raises the possibility that the language network may be relatively intact during sleep. Nevertheless, the effect of sleep on intrinsic language networks and whether sleep influences concordance with task-fMRI should be characterized in future studies.

Finally, our dataset represents a real-world sample of patients with brain tumors presenting for brain mapping. Brain tumors may cause reorganization of language networks depending on lesion location, recently demonstrated using task-fMRI [Wang et al., 2013]. In this study, we did not characterize the potential alterations in the organization of the language network, as our goal was to simply compare rs-fMRI to task-fMRI, and therefore we used each subject's task-fMRI as the target template to aid in selection of the best rs-fMRI language network. However, in using rs-fMRI as the sole source for language mapping, blinded rs-fMRI language network map selection therefore may be challenging depending on the degree to which the language network may be reorganized in specific subjects. Utilizing neural networks to classify resting-state networks [Mitchell et al., 2013] may be a promising approach although similar issues with classification of reorganized networks may continue to exist.

In conclusion, despite moderate overall concordance of rs-fMRI and task-fMRI language networks (and in some

individual cases, excellent concordance), the high subject level variability in the accuracy of rs-fMRI compared to task-fMRI warrants a view of optimistic caution in determining whether rs-fMRI has the potential to replace or supplement task-fMRI. iRMS may be used as a guide to determine the reliability of rs-fMRI for language network characterization. Improvements and standardization in rs-fMRI processing and analysis methods may allow for better delineation of language networks across subjects.

REFERENCES

Allen EA, Erhardt EB, Damaraju E, Gruner W, Segall JM, Silva RF, Havlicek M, Rachakonda S, Fries J, Kalyanam R, Michael AM, Caprihan A, Turner JA, Eichele T, Adelsheim S, Bryan AD, Bustillo J, Clark VP, Feldstein Ewing SW, Filbey F, Ford CC, Hutchison K, Jung RE, Kiehl KA, Kodituwakku P, Komesu YM, Mayer AR, Pearlson GD, Phillips JP, Sadek JR, Stevens M, Teuscher U, Thoma RJ, Calhoun VD (2011): A baseline for the multivariate comparison of resting-state networks. Front Syst Neurosci 5:2.

Bailey PD, Zacà D, Basha MM, Agarwal S, Gujar SK, Sair HI, Eng J, Pillai JJ (2015): Presurgical fMRI and DTI for the prediction of perioperative motor and language deficits in primary or metastatic brain lesions. J Neuroimaging 25:776–784.

Behzadi Y, Restom K, Liau J, Liu TT (2007): A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. NeuroImage 37:90–101.

Birn RM (2012): The role of physiological noise in resting-state functional connectivity. NeuroImage 62: 864–870.

Birn RM, Molloy EK, Patriat R, Parker T, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V (2013): The effect of scan length on the reliability of resting-state fMRI connectivity estimates. NeuroImage 83:550–558.

Bizzi A, Blasi V, Falini A, Ferroli P, Cadioli M, Danesi U, Aquino D, Marras C, Caldiroli D, Broggi G (2008): Presurgical functional MR imaging of language and motor functions: Validation with intraoperative electrocortical mapping. Radiology 248:579–589.

Castellanos FX, Di Martino A, Craddock RC, Mehta AD, Milham MP (2013): Clinical applications of the functional connectome. NeuroImage 80: 527–540.

Chang C, Glover GH (2010): Time–frequency dynamics of resting-state brain connectivity measured with fMRI. NeuroImage 50: 81–98.

de Chastelaine M, Rugg MD (2014): The relationship between task-related and subsequent memory effects. Hum Brain Mapp 35:3687–3700.

Cole DM, Smith SM, Beckmann CF (2010): Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. Front Syst Neurosci 4:8.

Damoiseaux JS, Rombouts SaRB, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF (2006): Consistent resting-state networks across healthy subjects. Proc Natl Acad Sci USA 103: 13848–13853.

Dym RJ, Burns J, Freeman K, Lipton ML (2011): Is functional MR imaging assessment of hemispheric language dominance as good as the wada test?: A meta-analysis. Radiology 261:446–455.

Fedorenko E, Thompson-Schill SL (2014): Reworking the language network. Trends Cogn Sci 18:120–126.

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Essen DCV, Raichle ME (2005): The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci USA 102:9673–9678.

Gabriel M, Brennan NP, Peck KK, Holodny AI (2014): Blood oxygen level dependent functional magnetic resonance imaging for presurgical planning. Neuroimaging Clin N Am 24:557–571.

Giussani C, Roux F-E, Ojemann J, Sganzerla EP, Pirillo D, Papagno C (2010): Is preoperative functional magnetic resonance imaging reliable for language areas mapping in brain tumor surgery? Review of language functional magnetic resonance imaging and direct cortical stimulation correlation studies [Review]. Neurosurgery 66:113–120.

Himberg J, Hyvärinen A, Esposito F (2004): Validating the independent components of neuroimaging time series via clustering and visualization. NeuroImage 22:1214–1222.

Hui M, Li J, Wen X, Yao L, Long Z (2011): An empirical comparison of information-theoretic criteria in estimating the number of independent components of fMRI data. PLoS One 6:e29274.

Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, Della Penna S, Duyn JH, Glover GH, Gonzalez-Castillo J, Handwerker DA, Keilholz S, Kiviniemi V, Leopold DA, de Pasquale F, Sporns O, Walter M, Chang C (2013): Dynamic functional connectivity: Promise, issues, and interpretations. NeuroImage 80:360–378.

Kalcher K, Huf W, Boubela RN, Filzmoser P, Pezawas L, Biswal B, Kasper S, Moser E, Windischberger C (2012): Fully exploratory network independent component analysis of the 1000 functional connectomes database. Front Hum Neurosci 6:301.

Kokkonen S-M, Nikkinen J, Remes J, Kantola J, Starck T, Haapea M, Tuominen J, Tervonen O, Kiviniemi V (2009): Preoperative localization of the sensorimotor area using independent component analysis of resting-state fMRI. Magn Reson Imaging 27:733–740.

Kollndorfer K, Fischmeister FPS, Kasprian G, Prayer D, Schöpf V (2013): A systematic investigation of the invariance of resting-state network patterns: Is resting-state fMRI ready for presurgical planning? Front Hum Neurosci 7:95.

Kristo G, Rutten G-J, Raemaekers M, de Gelder B, Rombouts SARB Ramsey NF (2014): Task and task-free FMRI reproducibility comparison for motor network identification. Hum Brain Mapp 35:340–352.

Kuchcinski G, Mellerio C, Pallud J, Dezamis E, Turc G, Rigaux-Viodé O, Malherbe C, Roca P, Leclerc X, Varlet P, Chrétien F, Devaux B, Meder J-F, Oppenheim C (2015): Three-tesla functional MR language mapping comparison with direct cortical stimulation in gliomas. Neurology 84:560–568.

Lee MH, Smyser CD, Shimony JS (2013): Resting-state fMRI: A review of methods and clinical applications. Am J Neuroradiol 34:1866–1872.

Lin Q-H, Liu J, Zheng Y-R, Liang H, Calhoun VD (2010): Semi-blind spatial ICA of fMRI using spatial constraints. Hum Brain Mapp 31:1076–1088.

Li Y-O, Adalı T, Calhoun VD (2007): Estimating the number of independent components for functional magnetic resonance imaging data. Hum Brain Mapp 28:1251–1266.

Mannfolk P, Nilsson M, Hansson H, Ståhlberg F, Fransson P, Weibull A, Svensson J, Wirestam R, Olsrud J (2011): Can resting-state functional MRI serve as a complement to task-based mapping of sensorimotor function? A test–retest reliability study in healthy volunteers. J Magn Reson Imaging 34:511–517.

Mazaika PK, Hoeft F, Glover GH, Reiss AL (2009): Methods and software for fMRI analysis of clinical subjects. NeuroImage 47(Suppl. 1):S58.

Mitchell TJ, Hacker CD, Breshears JD, Szrama NP, Sharma M, Bundy DT, Pahwa M, Corbetta M, Snyder AZ, Shimony JS, Leuthardt EC (2013): A novel data-driven approach to pre-operative mapping of functional cortex using resting-state functional magnetic resonance imaging. Neurosurgery 73:969–983.

Ogawa S, Menon RS, Tank DW, Kim SG, Merkle H, Ellermann JM, Ugurbil K (1993): Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. Biophys J 64:803–812.

Oldfield RC (1971): The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113.

Petrella JR, Shah LM, Harris KM, Friedman AH, George TM, Sampson JH, Pekala JS, Voyvodic JT (2006): Preoperative functional MR imaging localization of language and motor areas: Effect on therapeutic decision making in patients with potentially resectable brain tumors. Radiology 240:793–802.

Pillai JJ, Zaca D (2011): Relative utility for hemispheric lateralization of different clinical fMRI activation tasks within a comprehensive language paradigm battery in brain tumor patients as assessed by both threshold-dependent and threshold-independent analysis methods. NeuroImage 54(Suppl 1):S136–S145.

Pillai JJ (2010): The evolution of clinical functional imaging during the past 2 decades and its current impact on neurosurgical planning. Am J Neuroradiol 31:219–225.

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012): Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage 59:2142–2154.

Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE (2014): Methods to detect, characterize, and remove motion artifact in resting state fMRI. NeuroImage 84:320–341.

Qiu T, Yan C, Tang W, Wu J, Zhuang D, Yao C, Lu J, Zhu F, Mao Y, Zhou L (2014): Localizing hand motor area using resting-state fMRI: Validated with direct cortical stimulation. Acta Neurochir (Wien) 156:2295–2302.

R Core Team (2015): R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available at: http://www.R-project.org/.[AQ: Please provide the last accessed date for reference (R Core Team, 2015).]

Rosazza C, Aquino D, D'Incerti L, Cordella R, Andronache A, Zacà D, Bruzzone MG, Tringali G, Minati L (2014): Preoperative mapping of the sensorimotor cortex: comparative assessment of task-based and resting-State fMRI. PLoS One 9:e98860.

Rutten GJM, Ramsey NF, van Rijen PC, van Veelen CWM (2002): Reproducibility of fMRI-determined language lateralization in individual subjects. Brain Lang 80:421–437.

Sanai N, Mirzadeh Z, Berger MS (2008): Functional outcome after language mapping for glioma resection. N Engl J Med 358:18–27.

Shimony JS, Zhang D, Johnston JM, Fox MD, Roy A, Leuthardt EC (2009): Resting-state spontaneous fluctuations in brain activity: A new paradigm for presurgical planning using fMRI. Acad Radiol 16:578–583.

Tagliazucchi E, Laufs H (2014): Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. Neuron 82:695–708.

Tie Y, Rigolo L, Norton IH, Huang RY, Wu W, Orringer D, Mukundan S, Golby AJ (2014): Defining language networks from resting-state fMRI for surgical planning—A feasibility study. Hum Brain Mapp 35:1018–1030.

Van Dijk KRA, Sabuncu MR, Buckner RL (2012): The influence of head motion on intrinsic functional connectivity MRI. NeuroImage 59:431–438.

Voyvodic JT (2012): Reproducibility of single-subject fMRI language mapping with AMPLE normalization. J Magn Reson Imaging JMRI 36:569–580.

Wang L, Chen D, Yang X, Olson JJ, Gopinath K, Fan T, Mao H (2013): Group independent component analysis and functional MRI examination of changes in language areas associated with brain tumors at different locations. PLoS One 8:e59657

Wang Z, Liu J, Zhong N, Qin Y, Zhou H, Li K (2012): Changes in the brain intrinsic organization in both on-task state and post-task resting state. NeuroImage 62:394–407.

Wilke M, Holland SK, Ball WS (2003): Language processing during natural sleep in a 6-year-old boy as assessed by functional MRI. AJNR Am J Neuroradiol 24:42–44.

Xiang H-D, Fonteijn HM, Norris DG, Hagoort P (2010): Topographical functional connectivity pattern in the Perisylvian language networks. Cereb Cortex 20:549–560.

Zacà D, Jarso S, Pillai JJ (2013): Role of semantic paradigms for optimization of language mapping in clinical FMRI studies. AJNR Am J Neuroradiol 34:1966–1971.

Zacà D, Nickerson JP, Deib G, Pillai JJ (2012): Effectiveness of four different clinical fMRI paradigms for preoperative regional determination of language lateralization in patients with brain tumors. Neuroradiology 54:1015–1025.

Zhang D, Johnston JM, Fox MD, Leuthardt EC, Grubb RL, Chicoine MR, Smyth MD, Snyder AZ, Raichle ME, Shimony JS (2009): Preoperative sensorimotor mapping in brain tumor patients using spontaneous fluctuations in neuronal activity imaged with fMRI: Initial experience. Neurosurgery 65:226–236.