



Published in final edited form as:

Pain. 2016 March ; 157(3): 546–551. doi:10.1097/j.pain.0000000000000356.

Test-retest reliability of pain-related functional brain connectivity compared to pain self-report

Janelle E. Letzen, M.S.^a, Jeff Boissoneault, Ph.D.^{a,b}, Landrew S. Sevel, M.S.^a, and Michael E. Robinson, Ph.D.^a

^aDepartment of Clinical and Health Psychology, University of Florida, Gainesville, FL

^bPain Research and Intervention Center of Excellence, University of Florida, Gainesville, FL

Abstract

Test-retest reliability, or reproducibility of results over time, is poorly established for functional brain connectivity (fcMRI) during painful stimulation. As reliability informs the validity of research findings, it is imperative to examine, especially given recent emphasis on using functional neuroimaging as a tool for biomarker development. Although proposed pain neural signatures have been derived using complex, multivariate algorithms, even the reliability of less complex fcMRI findings has yet to be reported. The present study examined the test-retest reliability for fcMRI of pain-related brain regions, as well as self-reported pain [via visual analogue scales (VASs)]. Thirty-two healthy individuals completed three consecutive fMRI runs of a thermal pain task. Functional connectivity analyses were completed on pain-related brain regions. Intraclass correlations (ICC) were conducted on fcMRI values and VAS scores across the fMRI runs. ICC coefficients for fcMRI values varied widely (range = $-.174-.766$), with fcMRI between right nucleus accumbens and medial prefrontal cortex showing the highest reliability (range = $.649-.766$). ICC coefficients for VAS scores ranged from $.906-.947$. Overall, self-reported pain was more reliable than fcMRI data. These results highlight that fMRI findings might be less reliable than inherently assumed, and have implications for future studies proposing pain markers.

Keywords

Functional Connectivity; Test-Retest Reliability; Thermal Pain

Introduction

The psychometric properties of neuroimaging findings related to experimental, acute, and chronic pain are rarely reported. This knowledge is imperative, given the increased number of studies proposing neural markers of pain. These studies often inherently assume that neural markers are more reliable than self-reported pain [1,6,18]. In biomarker development

Corresponding Author: Michael E. Robinson, PhD, University of Florida, 101 South Newell Drive, Rm 3151, P.O. Box 100165, Gainesville, FL 32610-9165, Tel: 352-273-5220, Fax: 352-273-6156, merobin@ufl.edu.

There is no conflict of interest among authors.

across diseases, test-retest reliability (TRR) is a crucial measure, because unrepeatable results render findings **uninterpretable** [20].

TRR is the extent to which a dependent variable is consistent and error-free [15]. Importantly, TRR places an upper limit on validity. Whereas the TRR of self-reported pain has been extensively examined in clinical measures [12,16,25,36], there are presently no studies of pain neural markers that report information about TRR. In fact, very few studies have even examined TRR related simply to fMRI findings about pain processing.

We are aware of three studies examining the TRR of pain-related BOLD fMRI signal. These studies have used intraclass correlations (ICCs), which is a measure of TRR. Quiton et al. [26] measured intersession reliability of signal amplitude representing healthy individuals' brain activity during a thermal pain task. Across several days, ICCs ranged from 0.31–0.78 for task-related signal amplitude in “cortical pain network” regions. Similarly, we previously reported on the TRR of pain-related BOLD fMRI signal amplitude and cluster size in healthy controls undergoing a thermal pain task [17]. We compared these ICCs with that of participants' pain ratings captured via visual analogue scales (VASs). ICCs for fMRI values ranged from 0.32–0.88, whereas participants' pain ratings ranged from 0.93–0.96. Upadhyay et al. [34] had similar findings, with pain rating ICCs ranging from 0.86–0.94 for painful temperatures, and fMRI BOLD signal ranging from 0.5–0.85.

Because neural activity representing pain is likely better captured through coherent activation among pain-related regions, or functional connectivity (fcMRI) [29], the present study elaborates on previous work examining the TRR of pain-related BOLD fMRI by measuring TRR of fcMRI. Establishing the TRR of fcMRI findings is important because fcMRI analyses have been described as potentially “the best avenue to discover biomarkers of the brain in health and chronic pain” [6]. Additionally, correlation-based fcMRI can be used as the feature set entered into a machine learning model, which is a statistical technique that is commonly used in the development of neural markers for pain [28]. Our goal was not to propose or develop a new marker of pain; instead, the purpose of this study was to examine the TRR of fcMRI between pain-related brain regions of interest (ROIs) in a controlled, experimental environment. We also compared fcMRI TRR to that of self-reported pain, with a goal of determining whether this technology proved to be more reliable than pain ratings, in order to provide suggestions for future research on neural pain marker development.

Methods

The data used for this study were taken from the baseline visit of a larger, multi-visit study designed to examine mechanisms of placebo analgesia. During a screening visit for the larger study, thermal quantitative sensory testing (QST) was used to obtain individualized temperature thresholds producing a rating between 40–60 via VAS responses. This procedure was completed to make it more likely that participants received a stimulus during fMRI scanning that they would perceive as painful, as brain activity differs depending on whether a thermal stimulus is simply perceived as warm compared to painful [24]. Individuals who met study criteria then completed the first of three fMRI visits (i.e.,

baseline). Only thermal “pain” temperatures were used during the baseline visit, and no placebo conditioning or other manipulations were involved. Data used for the present study’s analyses were only taken from the baseline fMRI visit. Placebo conditioning was completed before either the second or third fMRI visits; however, those data are not reported in the present analyses.

Participants

Data from 32 pain-free participants (mean age = 22.5 y, SD = 3.2; 15 females) were analyzed to determine the TRR of VAS pain ratings and task-based fMRI. Twelve participants identified as Caucasian, 4 as African-American, 11 as Asian, 4 as Hispanic, and 1 as Pacific Islander. Exclusion criteria included: 1) enrollment in another research study that could influence participation; 2) inability to stop using pain medications seven days prior to undergoing QST; 3) history of psychological, neurologic, or psychiatric disorder; 4) history of chronic pain; 5) current medical condition that could interfere with study participation; 5) positive pregnancy test in females; 6) irremovable ferromagnetic metal within the body; and 7) inability to provide informed consent. The University of Florida Institutional Review Board approved the study, and all participants provided written informed consent to study procedures.

Experimental Materials

An MR-compatible, peltier-element-based stimulator (Medoc Thermal Sensory Analyzer, TSA-2001, Ramat Yishai, Israel) was used to deliver thermal stimuli during fMRI scanning. Temperatures ranged from 33°C to 51°C. Participants used computerized VASs (anchored by “no pain” and “the most intense pain imaginable”) to provide subjective pain ratings to thermal stimuli.

Experimental Procedures

Before participants completed the baseline fMRI visit, a screening visit was conducted to determine individualized temperatures. Location of the QST thermal pulses was on the dorsal aspect for both feet. Temperatures began at 43°C and increased by 1°C until tolerance or 51°C was reached. Pain intensity was rated on a VAS after each pulse. Thermal stimuli temperatures used during the baseline fMRI visit were set at each individual’s lowest temperature rated between 40–60 on the VAS. Participants were not told that only one temperature would be used during scanning procedures.

One anatomical and three task-based functional MRI scans were collected during the baseline fMRI visit. The task used for this study was consistent across all three functional scans, wherein 16 thermal stimuli were delivered to one of four sites on the dorsal aspects of both feet in a random order. Stimuli lasted for four seconds each, with a 12-second interstimulus interval. VAS pain intensity ratings were collected following each stimulus. Functional scans lasted five minutes and 40 seconds each. Runs were completed consecutively within the span of approximately thirty minutes.



Data Acquisition and Preprocessing

Scanning was completed on a 3.0T Phillips Achieva scanner, using an 8-channel head coil. A T1-weighted MP-RAGE protocol was used to collect a high-resolution, 3D, anatomical image. Protocol parameters included the following: 180 1mm sagittal slices, matrix (mm) = $256 \times 256 \times 180$, repetition time (TR) = 8.1ms, echo time (TE) = 3.7ms, FOV (mm) = $240 \times 240 \times 180$, FA = 8° , voxel size = 1mm^3). An echo planar acquisition protocol was used to collect functional images. Protocol parameters included: 38 contiguous 3mm trans-axial slices, matrix (mm) = $80 \times 80 \times 39$, TR/TE = 2000/30ms, FOV (mm) = $240 \times 240 \times 114$, FA = 80° , voxel size = 3mm^3). To reduce potential saturation effects from B_0 field inhomogeneity, four dummy volumes were discarded at the beginning of each functional scan.

SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK) in MATLAB 2011b (MathWorks, Sherbon, MA, USA) was used to preprocess fMRI data. Preprocessing procedures included 1) slice-time correction, 2) 3D motion correction with realignment to the middle volume of each sequence, 3) coregistration to the individual's structural MRI, 4) normalization to an MNI template, and 5) spatial smoothing [6mm^3 Gaussian kernel (FWHM)].

Motion parameters and signal-to-noise ratios (SNR) for all three fMRI runs were examined using a one-way ANOVA to determine whether systematic differences in these values might have influenced subsequent analyses. Average motion [$F(2,93) = 2.1, p > .05$] and average SNR [$F(2,93) = 2.297, p > .05$] were not significantly different among fMRI runs.

Functional Connectivity Analyses

To assess the fcMRI of *a priori* ROIs, we used the CONN toolbox [35] implemented through MATLAB. Each participant's preprocessed structural and functional images were entered into the toolbox. Confounding variables that affect fcMRI values were removed via CONN's CompCor algorithm for physiological noise [3], as well as temporal filtering and removal of confounding temporal covariates [35]. For task-based data, stimuli onsets and duration were specified in the toolbox, so that BOLD time series could be appropriately divided into task-specific blocks. Block regressors were then convolved with a canonical hemodynamic response function, and subsequently temporally filtered.

We then conducted 1) ROI-to-ROI analyses to determine fcMRI strength among selected ROIs and 2) seed-to-voxel analyses to measure fcMRI strength between each *a priori* ROI and all other voxels in the brain. ROIs used for this study were provided by default in the CONN toolbox.

For ROI-to-ROI analyses, we selected several *a priori* ROIs based on previous studies of pain processing. These ROIs included insular cortex (IC), medial prefrontal cortex (mPFC), anterior cingulate cortex (ACC), thalamus (Thal), and primary somatosensory cortex (S1). Within each ROI, the average BOLD time series is calculated across all voxels. Bivariate correlations were computed on each individual's time series from these ROIs for each fMRI run at the single-subject level. The bivariate correlations provide a measure of the linear association between each pair of the ROIs' averaged BOLD time series. As

standardized within the toolbox, correlations underwent a Fisher's Z-transformation to improve assumptions of normality [35]. ROI-to-ROI correlation matrices were yielded, and we extracted Z-transformed correlation values from *a priori* ROI pairs for each participant and run. These values were imported into SPSS v21.0 (SPSS Inc., Chicago, IL, USA) to calculate TRR.

Seed-to-voxel analyses were completed at the group-level to generate statistical connectivity maps for each *a priori* ROI. For these analyses, bivariate temporal correlations were calculated between the seed ROI and all other voxels in the brain, and subsequently Z-transformed. For a comprehensive explanation of CONN's processing stream, see Whitfield-Gabrieli and Nieto-Castanon [35].

Test-Retest Reliability

To measure TRR, we conducted two-way mixed single measures ICCs of consistency in SPSS v21.0 (SPSS Inc., Chicago, IL, USA) on Z-transformed correlation values from individual-level ROI-to-ROI matrices, as well as VAS pain ratings. ICCs are generally measured as the ratio of the variance of interest divided by the total variance (variance of interest and error; [32]). There are several versions of ICCs depending on the data characteristics to be measured [4]. Compared to ICCs of absolute agreement, ICCs of consistency assume that systematic differences among measures are irrelevant [23]. For the present study, the following formula was used in SPSS to calculate ICCs:

$$ICC(3,1) = \frac{(BMS - EMS)}{BMS + (k - 1)EMS}$$

In this equation, BMS refers to the between-subjects mean square (i.e., between-subject variability), whereas EMS refers to the error mean square (i.e., within-subject variability). The variable *k* is the number of scans per participant (i.e., 3 in the present study). In other words, a high ICC coefficient suggests that the within-subject variability across scans was relatively smaller than the between-subject variability across scans [33]. This ICC model has been used in previous studies to examine TRR of fMRI data [7,30,33]. For each ROI pair and pain ratings, we compared the following run pairs using ICCs: run 1 vs. run 2, run 2 vs. run 3, and run 1 vs. run 3. Additionally, we examined consistency across all three runs.

Results

Group-Level Functional Connectivity Analyses

All participants reported pain to thermal stimuli across runs (Table 1). Group-level ROI-to-ROI fMRI was significant for all ROI pairs when averaged across the three fMRI runs (Table 2), with right IC - right ACC showing the strongest average fMRI and right Thal - right IC showing the weakest fMRI. Individual runs also demonstrated significant group-level fMRI among ROIs, with the exception of the first run (right Thal - right IC fMRI was not significant). Figure 1 demonstrates group-level seed-to-voxel connectivity maps across runs with ROI-ROI overlays for ROI pairs tested in the right hemisphere.

Test-Retest Reliability of fcMRI and Pain Ratings

ICC coefficients range from 0 (no reliability) to 1 (perfect reliability). We classified ICC coefficients by the following descriptors based on previous work [8]: less than 0.4 = “poor,” between 0.4–0.6 = “fair,” between 0.61–0.8 = “good,” and greater than 0.8 = “excellent.” Results pertaining to ICC coefficients are discussed using these criteria.

Table 3 shows ICCs for ROI-ROI fcMRI across all three runs. The best ICC coefficient was for fcMRI between NAc and mPFC in the right hemisphere, which was in the “good” range for all fMRI runs. ICCs across runs for left NAc – mPFC and left IC – ACC fcMRI were between “fair” to “good” ranges. ICCs across runs were between the “poor” to “good” ranges for the following ROI pairs: right IC – ACC, left IC – left Thal, and bilateral Thal – respective ipsilateral S1. Of note, ICCs for the latter ROI pairs was highly variable between runs, with the poorest reliability comparing the first two runs. Table 4 shows ICC coefficients for participants’ pain ratings across all three fMRI runs. Comparing all runs, ICCs for self-reported pain all fell within the “excellent” range for TRR.

Discussion

One aim of this study was to measure the TRR of fcMRI among pain-related brain regions, because this information has not yet been reported. ICC coefficients were variable across ROI pairs, ranging from “poor” to “good.” The strongest and most consistent TRR was identified for right NAc-mPFC, whereas bilateral Thal-S1 and Thal-IC showed the poorest reliability.

Because some ROI pairs showed higher TRR, it is likely that these coactivations are better reproduced over time, and therefore more likely to be consistently interpretable. For example, only right NAc-mPFC showed “good” reliability among all runs. Functional connectivity between these regions has been previously described as a predictive “brain marker” for transition from acute to chronic pain [2], as well as involved in self-regulation for increasing and decreasing pain [37]. The results from the present study suggest that although fcMRI between these regions was not reproduced to a reliable enough degree in our study that would be appropriate to make individual-level conclusions, this fronto-striatal pathway warrants future research for understanding mechanisms of pain chronification and modulation. Comparatively, fcMRI between ROI pairs with poorer TRR (e.g., Thal-S1) should receive less focus until procedural or technological advances improve their reliability.

Reliability of fcMRI Compared to Self-Reported Pain

We also examined the TRR of pain intensity ratings collected via VASs to compare with the TRR of neuroimaging findings. ICCs among all run pairs for VAS pain ratings were in the “excellent” range of reliability. Overall, the pattern of ICCs in our study showed that fcMRI among the tested pain-related ROIs was not more reliable than participants’ VAS responses. These results suggest that fMRI findings do not necessarily exceed the reliability of individuals’ pain ratings, and the rationale of “flawed” self-report for biomarker development is potentially unfounded [1,6,18]. Future research examining the psychometric

properties of neural markers and how they compare to self-report will help establish the validity of this rationale.

Implications for the Development of Neural Pain Markers

As detailed in the National Pain Strategy, we agree that improved understanding of factors predicting the transition from acute to chronic pain is critical [22], as is the improvement of treatments and diagnostic techniques. Neuroimaging can be helpful in achieving these goals, mainly in making group-level conclusions. However, our results and others' suggest that reliability and validity of fMRI findings at the individual-level are questionable in their current states.

The most standard measure used to attain TRR is ICC [4]. ICCs for fcMRI in other disciplines have thus far proved widely disparate in range depending on the scan duration, preprocessing, modeling, and network tested [5,10,14,31]. Current estimates suggest that fcMRI ICCs range from 0.35–0.93, with ICCs of 0.5–0.6 most commonly reported. To provide a translational approach to neuroimaging findings, systematic work is needed to refine the psychometric properties of fMRI findings, especially for future studies reporting on neural pain markers. Because ICCs rely on between-subject variability to measure within-subject consistency, it should be noted that studies comparing a group with high between-subject variability to a group with low between-subject variability could result in artificial differences in ICC coefficients [4].

The present study was not aimed at proposing a neural marker of pain, given that the sensitivity and specificity of these ROI pairs for pain is not well established. Instead, we used a highly-controlled, experimental design as a “best case scenario” to measure TRR of fcMRI results, with the purpose of determining implications for future studies proposing neural pain markers. Machine learning techniques are commonly applied to neuroimaging marker development across clinical conditions. However, these techniques require feature extraction (i.e., selecting data as features that will be used in model training). Features are often based on data from general linear model analyses, such as regressions or correlations on the fMRI data [28]. Thus, measuring the TRR of these features is important, given that the TRR of the machine learning model is limited by the TRR of the data entered into the model.

Woo and Wager [38] suggested several desirable characteristics of neural pain markers. However, they did not specifically mention TRR as an imperative metric to be examined in biomarker development. Although classification metrics (e.g., sensitivity and specificity) are typically reported in these studies, TRR at the single-subject level is rarely reported, ultimately making the repeatability and validity of these markers unclear. Results from the present study suggest that TRR should be reported in future neural pain marker studies, especially as it compares to TRR of self-reported pain from study participants. Other suggestions for future research concern standardization in preprocessing and analysis procedures across studies to help establish the TRR and generalizability of findings. Future studies are needed to determine the optimal number of runs, or scan time, needed to produce the highest TRR.

Aside from focusing on TRR, we also suggest that future studies focus on validity and clinical utility of neural pain markers. Criteria for recovery and/or pain chronification vary widely in the literature, with various groups using disability [2,13], pain intensity [9], or return to work [11] as primary determinants. To date, studies regarding the use of neuroimaging markers have focused on reductions in pain intensity as their criteria for recovery [2,19]. It is not clear to what degree neural biomarkers identified to date reflect criteria that are ecologically or even statistically valid for a clinical setting (e.g., disability or work status).

Our study was the first report of fcMRI TRR among pain-related brain regions during pain processing. However, the study's limitations should be noted. Although valuable information was gained in highlighting that fcMRI results were as not as reliable as potentially presumed, we cannot make any conclusions about the reproducibility of neural pain markers that have already been proposed. Additionally, we acknowledge that our study design and sample were experimentally controlled to yield optimal reproducibility of fcMRI results and reported pain. This high degree of experimental control should have enhanced the possibility of finding highly reliable fcMRI results. In less controlled environments, it is possible that self-report would be less reliable than in an experimental setting. However, the same variables that would affect the reliability of pain ratings in these situations would presumably also affect associated brain activity, if fMRI were a valid tool for the assessment of pain. It should be noted that studies aimed at developing neural pain markers are also conducted in controlled experimental environments with stringent inclusion criteria.

It is still unclear how reproducible fcMRI results are within chronic pain populations, or how this TRR compares to that of patients' self-report. Although patients' accounts of pain in a clinic might vary due to other factors (e.g., psychological or motivational influences), again, it is likely that these factors would affect concomitant brain signatures. By definition, pain is always a subjective experience [21]. If a neural marker of pain is not altered with an individual's change in pain ratings, the neural marker is not likely explaining all of the variance associated with the individual's subjective experience of pain. Further, neural pain markers are typically validated against self-reported pain, which in turn acts as an upper limit for the psychometric properties of neuroimaging results [27].

Another potential concern for the development of neural pain markers is that patients presenting to a clinic with a primary complaint of pain are likely to have psychosocial and medical comorbidities. Future studies should examine effects of comorbid medical and psychological conditions, medications, and pain etiology on the TRR of fcMRI, as well as pain-specific neural markers in acute and chronic pain patients.

Conclusion

Functional neuroimaging is a very important tool to help elucidate mechanisms of pain etiology and treatment. However, TRR findings from the present study suggest that results should not be inherently assumed as highly reliable, and not necessarily more reliable than self-reported pain, as others have suggested [1,6,18]. Future work to improve psychometric properties of fMRI data will be essential for clinical translation of this tool.

Acknowledgements

This research was funded by the National Center for Complementary and Integrative Health through grants to Dr. Michael E. Robinson (R01AT001424-05A2) and Janelle E. Letzen (F31AT007898-01A2). Dr. Jeff Boissoneault was supported by a training grant from the National Institute of Neurological Disorders and Stroke (T32NS045551) to the University of Florida Pain Research and Intervention Center of Excellence.

References

1. Apkarian AV, Hashmi JA, Baliki MN. Pain and the brain: specificity and plasticity of the brain in clinical chronic pain. *Pain*. 2011; 152:S49–S64. [PubMed: 21146929]
2. Baliki MN, Petre B, Torbey S, Herrmann KM, Huang L, Schnitzer TJ, Fields HL, Apkarian AV. Corticostriatal functional connectivity predicts transition to chronic back pain. *Nat. Neurosci*. 2012; 15:1117–1119. [PubMed: 22751038]
3. Behzadi Y, Restom K, Liau J, Liu TT. A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI. *Neuroimage*. 2007; 37:90–101. [PubMed: 17560126]
4. Bennett CM, Miller MB. How reliable are the results from functional magnetic resonance imaging? *Ann. N. Y. Acad. Sci*. 2010; 1191:133–155. [PubMed: 20392279]
5. Birn RM, Molloy EK, Patriat R, Parker T, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V. The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage*. 2013; 83:550–558. [PubMed: 23747458]
6. Borsook D, Becerra L, Hargreaves R. Biomarkers for chronic pain and analgesia. Part 2: how, where, and what to look for using functional imaging. *Discov. Med*. 2011; 11:209. [PubMed: 21447280]
7. Braun U, Plichta MM, Esslinger C, Sauer C, Haddad L, Grimm O, Mier D, Mohnke S, Heinz A, Erk S. Test–retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *Neuroimage*. 2012; 59:1404–1412. [PubMed: 21888983]
8. Caceres A, Hall DL, Zelaya FO, Williams SCR, Mehta MA. Measuring fMRI reliability with the intra-class correlation coefficient. *Neuroimage*. 2009; 45:758–768. [PubMed: 19166942]
9. Campbell P, Wynne-Jones G, Muller S, Dunn KM. The influence of employment social support for risk and prognosis in nonspecific back pain: a systematic review and critical synthesis. *Int. Arch. Occup. Environ. Health*. 2013; 86:119–137. [PubMed: 22875173]
10. Fournier JC, Chase HW, Almeida J, Phillips ML. Model Specification and the Reliability of fMRI Results: Implications for Longitudinal Neuroimaging Studies in Psychiatry. *PLoS One*. 2014; 9:e105169. Available: <http://dx.doi.org/10.1371/journal.pone.0105169>. [PubMed: 25166022]
11. Gatchel RJ, Polatin PB, Mayer TG. The dominant role of psychosocial risk factors in the development of chronic low back pain disability. *Spine (Phila. Pa. 1976)*. 1995; 20:2702–2709. [PubMed: 8747248]
12. Grafton KV, Foster NE, Wright CC. Test-Retest Reliability of the Short-Form McGill Pain Questionnaire: Assessment of Intraclass Correlation Coefficients and Limits of Agreement in Patients With Osteoarthritis. *Clin. J. Pain*. 2005; 21 Available: http://journals.lww.com/clinicalpain/Fulltext/2005/01000/Test_Retest_Reliability_of_the_Short_Form_McGill.9.aspx.
13. Grotle M, Brox JI, Glomsrød B, Lønn JH, Vøllestad NK. Prognostic factors in first-time care seekers due to acute low back pain. *Eur. J. Pain*. 2007; 11:290–298. [PubMed: 16677837]
14. Jann K, Gee DG, Kilroy E, Schwab S, Smith RX, Cannon TD, Wang DJJ. Functional connectivity in BOLD and CBF data: similarity and reliability of resting brain networks. *Neuroimage*. 2015; 106:111–122. [PubMed: 25463468]
15. Jensen MP. The validity and reliability of pain measures in adults with cancer. *J. Pain*. 2003; 4:2–21. [PubMed: 14622723]
16. Lacourt TE, Houtveen JH, van Doornen LJP. Experimental pressure-pain assessments: Test–retest reliability, convergence and dimensionality. *Scand. J. Pain*. 2012; 3:31–37.
17. Letzen JE, Sevel LS, Gay CW, O’Shea AM, Craggs JG, Price DD, Robinson ME. Test-Retest Reliability of Pain-Related Brain Activity in Healthy Controls Undergoing Experimental Thermal Pain. *J. Pain*. 2014; 15:1008–1014. [PubMed: 24998897]

18. Lu Y, Klein GT, Wang MY. Can pain be measured objectively? *Neurosurgery*. 2013; 73:24–25.
19. Mansour A, Baliki MN, Huang L, Torbey S, Herrmann K, Schnitzer TJ, Apkarian AV. Brain white matter structural properties predict transition to chronic pain. *Pain*. 2013; 154:2160–2168. [PubMed: 24040975]
20. Mayeux R. Biomarkers: potential uses and limitations. *NeuroRx*. 2004; 1:182–188. [PubMed: 15717018]
21. Merskey H, Bogduk N. IASP Task Force on Taxonomy Part III: Pain Terms, A Current List with Definitions and Notes on Usage. IASP Task Force Taxon. 1994:209–214. Available: <http://www.iasp-pain.org/Content/NavigationMenu/GeneralResourceLinks/PainDefinitions/default.htm#Pain>.
22. National Pain Strategy. 2015
23. Nichols DP. Choosing an intraclass correlation coefficient. *SPSS keywords*. 1998; 67:1–2.
24. Peltz E, Seifert F, DeCol R, Dörfler A, Schwab S, Maihöfner C. Functional connectivity of the human insular cortex during noxious and innocuous thermal stimulation. *Neuroimage*. 2011; 54:1324–1335. doi:<http://dx.doi.org/10.1016/j.neuroimage.2010.09.012>. [PubMed: 20851770]
25. Price DD, Patel R, Robinson ME, Staud R. Characteristics of electronic visual analogue and numerical scales for ratings of experimental pain in healthy subjects and fibromyalgia patients. *Pain*. 2008; 140
26. Quiton RL, Keaser ML, Zhuo J, Gullapalli RP, Greenspan JD. Intersession reliability of fMRI activation for heat pain and motor tasks. *NeuroImage Clin*. 2014; 5:309–321. [PubMed: 25161897]
27. Robinson ME, Staud R, Price DD. Pain measurement and brain activity: will neuroimages replace pain ratings? *J. Pain*. 2013; 14:323–327. [PubMed: 23548484]
28. Rosa MJ, Seymour B. Decoding the matrix: Benefits and limitations of applying machine learning algorithms to pain neuroimaging. *Pain*. 2014; 155:864–867. [PubMed: 24569148]
29. Shackman AJ, Salomons TV, Slagter HA, Fox AS, Winter JJ, Davidson RJ. The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat Rev Neurosci*. 2011; 12:154–167. Available: <http://dx.doi.org/10.1038/nrn2994>. [PubMed: 21331082]
30. Shehzad Z, Kelly AMC, Reiss PT, Gee DG, Gotimer K, Uddin LQ, Lee SH, Margulies DS, Roy AK, Biswal BB, Petkova E, Castellanos FX, Milham MP. The Resting Brain: Unconstrained yet Reliable. *Cereb. Cortex*. 2009; 19:2209–2229. [PubMed: 19221144]
31. Shirer WR, Jiang H, Price CM, Ng B, Greicius MD. Optimization of rs-fMRI Pre-processing for Enhanced Signal-Noise Separation, Test-Retest Reliability, and Group Discrimination. *Neuroimage*. 2015; 117:67–79. [PubMed: 25987368]
32. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull*. 1979; 86:420. [PubMed: 18839484]
33. Song J, Desphande AS, Meier TB, Tudorascu DL, Vergun S, Nair VA, Biswal BB, Meyerand ME, Birn RM, Bellec P. Age-related differences in test-retest reliability in resting-state brain functional connectivity. *PLoS One*. 2012; 7:e49847. [PubMed: 23227153]
34. Upadhyay J, Lemme J, Anderson J, Bleakman D, Large T, Evelhoch JL, Hargreaves R, Borsook D, Becerra L. Test-retest reliability of evoked heat stimulation BOLD fMRI. *J. Neurosci. Methods*. 2015:1–9.
35. Whitfield-Gabrieli S, Nieto-Castanon A. : A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. *Brain Connect*. 2012; 2:125–141. [PubMed: 22642651]
36. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. *J. Clin. Nurs*. 14:798–804. n. d.; Available: <http://cat.inist.fr/?aModele=afficheN&cpsid=16921911>. [PubMed: 16000093]
37. Woo C-W, Roy M, Buhle JT, Wager TD. Distinct Brain Systems Mediate the Effects of Nociceptive Input and Self-Regulation on Pain. *PLoS Biol*. 2015; 13:e1002036. Available: <http://dx.doi.org/10.1371/journal.pbio.1002036>. [PubMed: 25562688]
38. Woo C-W, Wager TD. Neuroimaging-based biomarker discovery and validation. *Pain*. 2015; 156 Available: http://journals.lww.com/pain/Fulltext/2015/08000/Neuroimaging_based_biomarker_discovery_and.5.aspx.

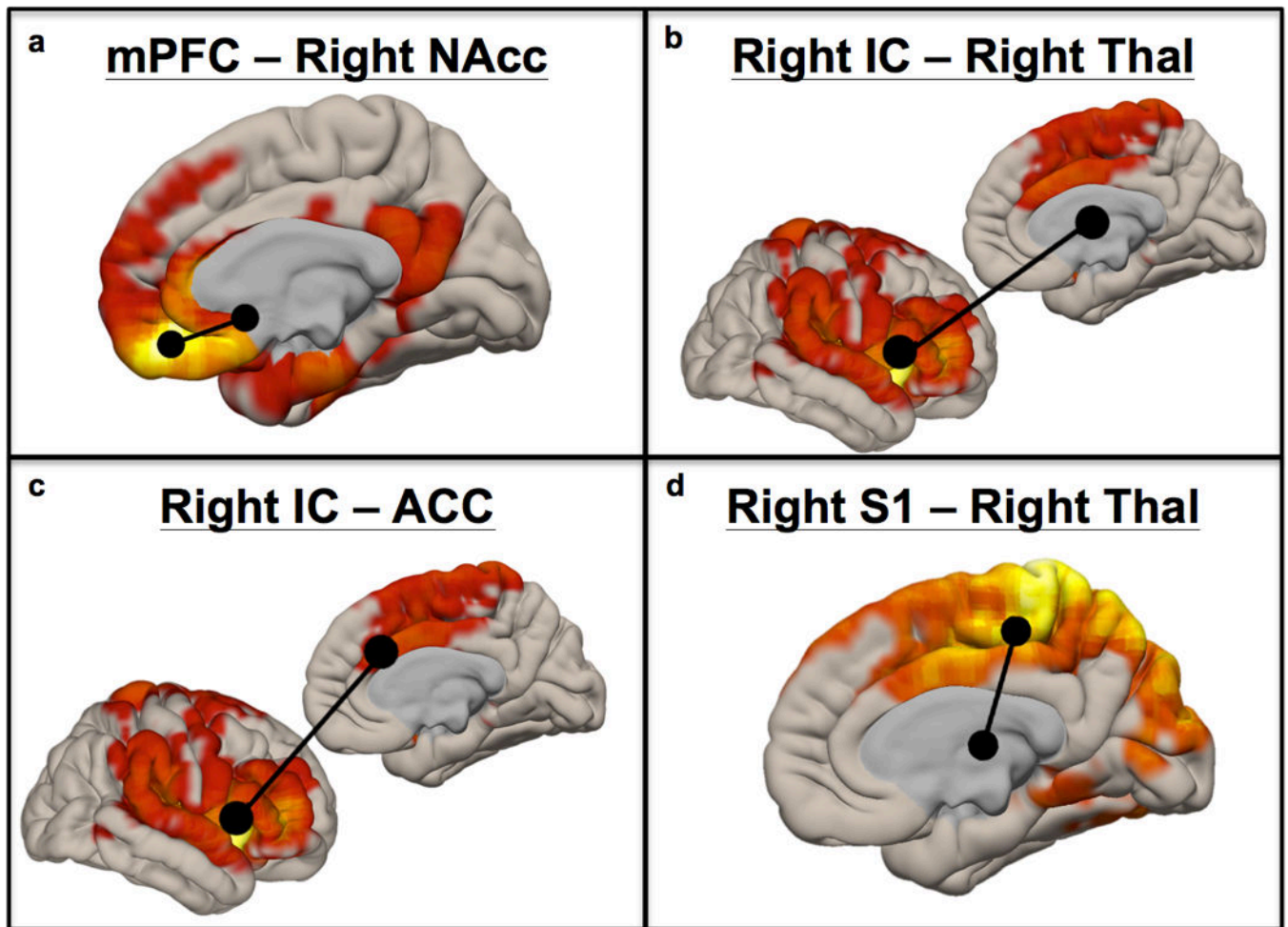


Figure 1.

Cortical renderings of group-level fcMRI between *a priori* ROIs and the whole brain for voxels significantly correlated at $p_{FWE} < .001$. Figure 1a used mPFC as a seed region, 1b and 1c used right insular cortex, and 1d used right posterior gyrus (S1). Areas used for ROI-to-ROI fcMRI analyses are shown in black over each statistical map.

Table 1

Group-level descriptive statistics for subjective VAS pain ratings across all three fMRI runs.

Run	Mean	SD
Average	41.75	13.97
1	42.91	12.88
2	41.43	15.70
3	40.91	15.30

Table 2

Group-level functional connectivity statistics for ROIs among fMRI runs.

ROI-ROI	Run	beta	T
R NAc – mPFC	Average	.35	6.92***
	1	.38	6.07***
	2	.39	6.49***
	3	.30	4.78***
L NAc – mPFC	Average	.35	6.92***
	1	.35	6.55***
	2	.29	4.82***
	3	.33	5.77***
R Thal – R IC	Average	.19	5.07***
	1	.12	1.99
	2	.24	3.85**
	3	.21	3.79**
L Thal – L IC	Average	.24	5.29***
	1	.24	3.92**
	2	.24	4.43***
	3	.23	4.18***
R IC – ACC	Average	.36	7.95***
	1	.27	4.48***
	2	.37	6.52***
	3	.43	6.92***
L IC – ACC	Average	.30	6.27***
	1	.25	3.62**
	2	.34	6.17***
	3	.32	5.20***
R Thal – R S1	Average	.21	5.61***
	1	.21	3.69**
	2	.21	3.65**
	3	.22	3.90**
L Thal – L S1	Average	.22	6.39***
	1	.21	3.93**
	2	.19	4.22***
	3	.24	5.05***

* significant at $p < .05$;

** significant at $p < .01$;

*** significant at $p < .001$

Region of interest (ROI), right (R), left (L), nucleus accumbens (NAc), medial prefrontal cortex (mPFC), thalamus (Thal), anterior cingulate cortex (ACC), insular cortex (IC), primary somatosensory cortex (S1)

Table 3

Intraclass correlation coefficients for fcMRI between *a priori* ROIs among fMRI runs.

ROI-ROI	Run 1 vs. Run 2	Run 2 vs. Run 3	Run 1 vs. Run 3	All Runs
R NAc – mPFC	.700**	.649**	.706***	.766***
L NAc – mPFC	.509*	.737***	.425	.671***
R Thal – R IC	.349	.250	–.001	.294
L Thal – L IC	.258	.418	.461*	.482*
R IC – ACC	.532*	.474*	.364	.557**
L IC – ACC	.518*	.596**	.614**	.672***
R Thal – R S1	–.006	.641**	.230	.422*
L Thal – L S1	–.174	.591**	.180	.328

* significant at $p < .05$;

** significant at $p < .01$;

*** significant at $p < .001$

Region of interest (ROI), right (R), left (L), nucleus accumbens (NAc), medial prefrontal cortex (mPFC), thalamus (Thal), anterior cingulate cortex (ACC), insular cortex (IC), primary somatosensory cortex (S1)

Table 4

Intraclass correlation coefficients for VAS pain ratings test-retest reliability among fMRI runs

Run 1 vs. Run 2	Run 2 vs. Run 3	Run 1 vs. Run 3	All Runs
.936***	.925***	.906***	.947***

* significant at $p < .05$;** significant at $p < .01$;*** significant at $p < .001$