HHS Public Access

Author manuscript

Pain. Author manuscript; available in PMC 2022 April 01.

Published in final edited form as:

Pain. 2021 April 01; 162(4): 1241–1249. doi:10.1097/j.pain.0000000000002108.

Thalamic neuroinflammation as a reproducible and discriminating signature for chronic low back pain

Angel Torrado-Carvajal[1,2], **Nicola Toschi**[1,3], **Daniel S Albrecht**[1], **Ken Chang**[1], **Oluwaseun Akeju**[4], **Minhae Kim**[1], **Robert R Edwards**[5], **Yi Zhang**[4], **Jacob M Hooker**[1], **Andrea Duggento**[1,3], **Jayashree Kalpathy-Cramer**[1], **Vitaly Napadow**[1], **Marco L Loggia**[1,#]

[1]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

[2]Medical Image Analysis and Biometry Laboratory, Universidad Rey Juan Carlos, Madrid, Spain.

[3]Department of Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy.

[4]Department of Anesthesia, Critical Care and Pain Medicine, MGH/HMS, Boston, MA, USA.

[5]Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, HMS, Boston, MA, USA.

Abstract

Using positron emission tomography, we recently demonstrated elevated brain levels of the 18kDa translocator protein (TSPO), a glial activation marker, in chronic low back pain (cLBP) patients, compared to healthy controls (HC). Here, we first sought to replicate the original findings in an independent cohort (15 cLBP, 37.8±12.5 y/o; 18 HC, 48.2±12.8 y/o). We then trained random forest (RF) machine learning algorithms based on TSPO imaging features combining discovery and replication cohorts (totaling 25 cLBP, 42.4±13.2 y/o; 27 HC, 48.9±12.6 y/o), in order to explore whether image features other than the mean contain meaningful information that might contribute to the discrimination of cLBP patients and HC. Feature importance was ranked usind SHapley Additive exPlanations (SHAP) values, and the classification performance (in terms of AUC values) of classifiers containing only the mean, other features, or all features was compared using the DeLong test. Both region-of-interest (ROI) and voxelwise analyses replicated the original observation of thalamic TSPO signal elevations in cLBP patients compared to HC ($p's<0.05$). The RF-based analyses revealed that while the mean is a discriminating feature, other features demonstrate similar level of importance, including the maximum, kurtosis and entropy.

Our observations suggest that thalamic neuroinflammatory signal is a reproducible and discriminating feature for cLBP, further supporting a role for glial activation in human chronic low back pain, and the exploration of neuroinflammation as a therapeutic target for chronic pain. This work further shows that TSPO signal contains a richness of information that the simple mean might fail to capture completely.

Corresponding Author: Marco L Loggia, PhD, marco.loggia@mgh.harvard.edu.

Conflict of interest statement: J.K. is a consultant/advisory board member for Infotech, Soft. The other authors declare no competing financial interests.

Keywords

chronic low back pain; classification; neuroinflammation; PET/MR

1. INTRODUCTION

Recently, we showed that patients with chronic low back pain (cLBP) demonstrate an increased brain uptake of [^{11}C]-PBR28 [32], a radioligand that binds to the 18kDa translocator protein (TSPO) [12; 13; 24]. While TSPO is a mithocondrial protein ubiquitously found in multiple cell types throughout the body, it is widely used as a marker of glial activation because its expression levels are low in the healthy central nervous system, but dramatically upregulated during neuroinflammatory responses and highly colocalized with activated microglia/macrophages and astrocytes [41]. As such, the elevation in [^{11}C]-PBR28 uptake we reported in cLBP patients suggests that neuroinflammation may play a role in the pathophysiology of human chronic pain. Indeed, following this initial observation, we reported [^{11}C]-PBR28 elevations also in the brain of patients suffering from fibromyalgia [3] and migraine [5], as well as in the spinal cord and dorsal root ganglia / nerve roots of patients with lumbar radiculopathy [2]. Using magnetic resonance spectroscopy, we have also shown that patients with fibromyalgia demonstrate elevated brain levels of choline, a potential marker of astrogliosis [28]. Taken together, these results suggest that glial activation/neuroinflammation does indeed occur in various human chronic pain disorders, and that it can be observed across multiple levels of the nervous system, in accordance with many animal studies (e.g., [14; 45; 49]).

Confirmation of elevated glial markers in chronic pain patients might have several clinical implications, including by providing support for a more aggressive exploration of neuroinflammation as a therapeutic target. Furthermore, the fact that the spatial distribution of the elevated TSPO signal is very different across subtypes of chronic pain patients (e.g., most consistently involving the thalamus in cLBP, more widespread in the cortex seemingly without substantial subcortical involvement in fibromyalgia) raises the possibility that different “neuroinflammatory signatures” might be idenfitied in different conditions characterized by persistent pain.

The overall aims of this project were twofold. First, we evaluated whether thalamic neuroinflammation in chronic low back pain, which we initially described in a small study [32], was a reliable and reproducible observation. To this end, we first sought to replicate the findings of our previous study (i.e., the discovery dataset) demonstrating that [^{11}C]-PBR28 brain uptake increases in a new, independent cohort of cLBP patients (i.e., the replication dataset). The second aim was to explore the richness of the TSPO signal, and specifically the ability of its various features, in addition to the mean, to discriminate cLBP patients from controls. To address this aim, we employed a radiomics approach to extract various image features (e.g., maximum, kurtosis, entropy) from the full dataset (i.e. the combination of discovery and replication datasets), and then used novel methods for the ranking of feature importance based on a random forest classifier. Finally, we tested the classification performance of models containing only the mean, only features other than the mean, or all

features. These approaches allowed us to evaluate whether the neuroinflammatory signal contains a richness of information that the simple mean might fail to completely capture.

2. MATERIALS AND METHODS

This study was conducted at the Athinoula A. Martinos Center for Biomedical Imaging at Massachusetts General Hospital. The protocol was approved by the local Institutional Review Board and the Radioactive Drug Research Committee.

2.1. Participants

Replication study.—Thirty one patients with cLBP, defined as ongoing and persistent pain for more than 6 months (with or without a radicular component), and forty five healthy volunteers (HC) with no history of chronic pain were screened for participation in this study. Participants were excluded if they had any PET/MR contraindications such as metallic implants, history of head trauma or claustrophobia, history of major medical or psychological disorders, were on benzodiazepines or anticoagulants, or had comorbid pain disorders more severe than low back pain. Because the Ala147Thr polymorphism in the *TSPO* gene predicts binding affinity for [^{11}C]-PBR28, with the Ala/Ala, Ala/Thr, and Thr/Thr genotypes being associated with high, mixed, and low affinity binding, respectively [29; 40; 52], all participants were tested for this polymorphism and those with Thr/Thr genotype were excluded from the study. In total, 15 cLBP (8 female; 37.5±12.5 y/o) and 18 HC (10 female; 47.9±12.9 y/o) participants were found eligible and underwent [^{11}C]-PBR28 PET/MR brain imaging scans, forming the replication dataset. Of these participants, 22 were Ala/Ala (i.e., high-affinity binders; HABs: cLBP n=12, HC n=10), and 11 were Ala/Thr (i.e., mixed-affinity binders; MABs: cLBP n=3, HC n=8).

Imaging feature assessment.—In addition to the data from the patients and controls newly recruited for the replication study, data from the original [^{11}C]-PBR28 cLBP discovery study [32] were included for the purpose of exploring whether TSPO features other than the mean contain meaningful information that might contribute to the discrimination of cLBP patients from controls. This discovery dataset included 10 additional patients (5 female; 48.9±11.6 y/o) and 9 additional controls (4 female; 50.1±12.4). Of these participants, 15 were Ala/Ala (HABs: cLBP n=8, HC n=7), and 4 were Ala/Thr (MABs: cLBP n=2, HC n=2). For more details, see [32]. By pooling the discovery and replication datasets, a total of 52 datasets [i.e., 25 patients (13 female; 20 HABs; 42.4±13.2 y/o) and 27 controls (14 female; 17 HABs; 48.9±12.6 y/o)] were included in the final sample used for cLBP discrimination. Of note, the healthy volunteers in the replication dataset were obtained from an existing dataset of participants recruited through multiple different studies, and thus not specifically recruited to match the newly recruited patients in the replication dataset. As a result, the former group happened to be marginally older (p=0.03; two-sided two-sample t-test). Nonetheless, no statistically significant differences in age were observed when considering all available cLBP and control data (p=0.07). Supplementary Table 1 includes detailed demographics of the discovery and replication cohorts.

2.2. Data Acquisition

Eligible participants were invited to participate in an imaging visit. At the beginning of the visit, venous blood was drawn to assess levels of the circulating proinflammatory cytokine IL-6. PET and MR data were acquired simultaneously on a Siemens PET/MRI scanner [15] consisting of a dedicated brain avalanche photodiode-based PET scanner operating in the bore of a 3T whole-body MR scanner (Siemens Healthineers, Erlangen, Germany). PET emission data were acquired in 3-dimensional mode for 90 minutes starting with the administration of [^{11}C]-PBR28 injected intravenously with a slow bolus over 30s (see supplementary Table 1). MR imaging included a T1-weighted volume (MEMPRAGE; TR/TE1/TE2/TE3/TE4= 2530/1.64/3.5/5.36/7.22ms, flip angle=7°, voxel size=1×1×1mm, acquisition matrix=280×280×208). During the imaging visit, all subjects rated their level of pain using a 0–100 numerical ratings scale (pain: 0 = 'no pain', 100 = 'most intense pain tolerable'). The participants also completed the McGill Pain Questionnaire (MPQ), short form [38].

2.3. Data Analysis

Replication study.—As in the original study, standardized uptake values (SUV) from [^{11}C]-PBR28 data collected 60–90 min post-injection images were normalized to MNI space using non-linear registration (FSL FNIRT) [25]. SUV images were spatially smoothed (full-width at half-maximum=8mm) as well as whole-brain normalized in order to account for global signal differences across subjects (SUVR). Of note, measures obtained with kinetic modeling using arterial input functions (e.g., distribution volume [V_T] and V_T ratio) are still largely considered the gold standard for quantification of TSPO tracers. However, we have previously shown that in cLBP patients SUVR is highly correlated to V_T ratio [4; 6]. Moreover, the use of a simplified metric based on a ratio was found to increase the sensitivity to detect neuroinflammatory responses in various conditions [6; 35; 46]. Indeed, our group and others have used ratio metrics, including SUVR, to demonstrate higher TSPO signal across multiple conditions (compared to controls), in spatial distributions overlapping with the known or expected distribution of neuroinflammation in each condition (e.g., basal ganglia in Huntington's Disease [33], temporoparietal regions in Alzheimer's disease [35], motor/premotor cortices and corticospinal tracts in amyotrophic lateral sclerosis [7; 53]).

In the discovery dataset, cLBP patients demonstrated higher [^{11}C]-PBR28 SUVR in the thalamus, both in region-of-interest (ROI) as well as voxelwise analyses. Thus, in the present study, our primary analyses were centered on the thalamus, which we interrogated using both an ROI-based approach and a thalamus-focused voxelwise approach. To this end, left and right thalamic ROIs were computed by thresholding probabilistic labels from the Harvard-Oxford Atlas, using an arbitrary threshold of 0.3. Mean [^{11}C]-PBR28 SUVR within these ROIs was computed for all participants, and compared across groups using an analysis of variance (ANOVA) including hemisphere (left, right) as within-subject factor, as well as genotype/binding affinity status (HAB, MAB) and age as nuisance covariates. The model also included a group*hemisphere interaction term. Subsequently, voxelwise group differences in thalamic [^{11}C]-PBR28 SUVR were evaluated using non-parametric voxelwise permutation analyses (FSL randomize, n=5000 permutations, enhanced using threshold-free cluster enhancement; [51]) and corrected for multiple comparisons, again including

genotype and age as covariates. The primary voxelwise analyses were performed by restricting the search area to the left and right thalamic ROIs to be maximally sensitive to localized changes in these a priori chosen regions. Moreover, because in the original study the cLBP patients also demonstrated additional significant clusters of elevated [^{11}C]-PBR28 SUVR in the cortex (primary somatosensory and motor cortices), the voxelwise analyses were repeated using the same methods, but over the entire brain. The anatomical location of the clusters showing a statistical difference between cLBP and controls was assessed using the Morel stereotactic atlas of the human thalamus [39]. Pearson's correlation analyses were used to evaluate the relationship between pain ratings, MPQ scores, IL-6 levels, and genotype corrected [^{11}C]-PBR28 SUVR, either as mean signal extracted from the thalamus, or as the number of voxels with an SUVR≥1.4 [32].

Imaging feature assessment.—From the pooled dataset (discovery and replication) and for each of the two thalamic ROIs, we computed 10 radiomic features based on standard within-ROI distribution statistics of the [^{11}C]-PBR28 SUVR (i.e., mean, minimum, maximum, range, standard deviation, energy, entropy, kurtosis, skewness, covariance). Radiomics, so far mostly used in oncology, uses data-characterization algorithms to extract large amounts of quantitative features, potentially uncovering characteristics that may be invisible to the naked eye [21]. Given the relatively small size of our final sample, the number of radiomic features extracted was purposefully kept relatively small, in order to prevent overfitting [44]. Furthermore, the focus on thalamic ROIs was based on the fact that thalamic, but not cortical, TSPO signal elevations were observed in the replication dataset (see below). Note that even if only the left thalamus demonstrated significant TSPO signal elevations in voxelwise analyses, similar group differences were observed in both thalami, and no group*hemisphere interaction was statistically significant in ROI analyses (see below). Thus, in our evaluation of the radiomic features we included both left and right thalami.

We used a genetic search pipeline, which uses an evolutionary algorithm to explore hyperparameter values as well as classifer architectures and their combination [31] in a nested cross-validation fashion. The search pipeline selected an optimized random forest (RF) classifier which was implemented using the scikit-learn python module [42]. RF is an ensemble learning algorithm that combines a number of decision trees into a single predictive algorithm [11; 19]. Decision tree classifiers are trained on various sub-samples of the cohort and results are averaged to improve the predictive performance. Moreover, due to the high number of decision trees trained on bootstrapped subsets of the training data, the RF method is resistant to overfitting.

For the current study, we ran occlusion tests to assess whether hemisphere (left, right or bilateral thalamus), as well as feature type (mean, other features, or all features) had an impact on the classification performance. We used a repeated stratified 5-fold cross validation approach, with 1000 repetitions (different randomization in each repetition) to assess each classifier. Cross-validation is a procedure to assess predictive models by splitting the original cohort into a training and a test set, to independently create a model with some of the samples and validate it over the remaining ones. In k-fold cross-validation, the original cohort is partitioned into k equal size random subsets. Of the k subsets, k-1 are used

as training data and the remaining one is used as the validation set for testing the model. This process is repeated k times, with each of the k subsets used once as the validation set. The k results obtained after k-fold cross-validation are then averaged to produce an overall estimate. In classification, it is common to use stratified k-fold cross-validation, in which the cohort is split so that each fold maintains the same proportions of classes.

The performance of each classifier was assessed by calculating the area under the curve (AUC) from receiver operating characteristic (ROC) curve analysis for each one of the k splits generated during cross-validation. Median accuracy, sensitivity, and specificity were also computed. SHapley Additive exPlanations (SHAP) values were computed to explore the radiomic features contribution to the model output. This method is agnostic (i.e., model-independent), was designed to be unambiguous from a mathematical point of view and has never before been applied to a pain imaging dataset. SHAP values unify six existing methods for quantifying feature importance providing results that are more consistent with human intuition as compared with previous approaches [34]. ROC curves from different classifiers obtained in the occlusion tests were compared for AUC differences using a paired DeLong test [20], and resulting *p*-values were corrected for multiple comparisons using Bonferroni correction. The same analyses were performed after correcting for genotype and age, for consistency with the analyses detailed in the replication study section (see Supplementary Document 1).

3. RESULTS

Replication study.

Analysis of mean [^{11}C]-PBR28 SUVR in the thalamic ROIs demonstrated a significantly higher signal in patients with cLBP compared to controls (group effect: $F(1,29)=4.51$, $p<0.05$; Figure 1). The group*hemisphere interaction was not statistically significant ($F(1,29)=0.69$, $p = 0.41$, n.s.), indicating that the group effect was not significantly different across hemispheres. The voxel-wise distribution of thalamic SUVRs (Supplementary Figure 1) revealed that in control subjects non-zero median voxel counts were observed only below values of 1.5, whereas in patients with chronic LBP a substantial number of voxels demonstrated values higher than 1.5 in both hemispheres (medians: 25 and 45 in the left and right thalamus, respectively). Additional analyses of the remaining radiomic features in the thalamic ROIs could be found in Supplementary Document 1.

In thalamus-focused voxel-wise analyses (Figure 2), patients demonstrated a statistically significant [^{11}C]-PBR28 SUVR elevation in one cluster (cluster size = 263 voxels) showing two local maxima in the left thalamus, compared to controls. The first local maximum was located in a ventral portion (peak: x = -12mm, y = -22mm, z = 0mm; Z-stat = 3.80) consistent with the ventral posterior lateral, medial and inferior (VPL, VPM, VPI), ventral lateral posterior (VLP), anterior pulvinar (PuA), posterior (Po), lateral posterior (LP), and centromedian (CM) nuclei. This peak was relatively close to the thalamic region demonstrating statistically significant elevation in PET signal in the original study, but was located slightly more ventrally, laterally and posteriorly. The second local maximum (peak: x = -16mm, y = -20mm, z = +14mm; Z-stat = 3.31) was in a position consistent with a more dorsal portion of the VLP and LP nuclei. As in the original study, no right thalamic clusters

reached statistical significance in voxelwise group analyses. Also, no cortical group effects were observed in whole-brain analyses, and no brain regions showed statistically higher PET signal in the controls compared to the patients, even when lowering the threshold to a very lenient p=0.1. Finally, no significant associations were observed between any PET measures investigated (i.e., mean [^{11}C]-PBR28 SUVR from thalamic ROIs or significant thalamic clusters, and number of voxels with a SUVR higher than 1.4) and pain ratings or serum IL-6 levels (p's>0.08).

Imaging feature assessment.

Figure 3 shows the median (across folds) ROC curves for the classifiers selected by the genetic search based on each subset of features in the occlusion tests. A direct pairwise comparison across ROC curves revealed that classifiers including features other than mean [^{11}C]-PBR28 thalamic signal contained additional information when compared to those including only the mean (p's < 0.001), but similar information when compared to those including all features, including the mean (p's ≥ 0.07; Table 1). For instance, when using features from both thalami, the AUC went from 0.79±0.10 (median ± MAD) when using the mean values only to 0.88±0.07 when using all other features except for the mean, but did not increase further when adding back the mean value into the feature set (0.88±0.07). Predictions from the RF classifiers showed good classification capabilities (better than random, i.e., AUC = 0.88±0.07) between cLBP patients and healthy controls, demonstrating a median accuracy, sensitivity and specificity of 0.70±0.10, 0.80±0.20, and 0.80±0.20, respectively. After correcting for age and genotype, classifiers including features other than mean [^{11}C]-PBR28 thalamic signal again demonstrated to contain additional information when compared to those including only the mean (p's < 0.001; Supplementary Table 2), but similar information when compared to those including all features together, including the mean (p's = 1). For instance, when using age- and genotype-corrected features from both thalami, the AUC went from 0.69±0.11 when using the mean values only to 0.80±0.10 when using all other features, but did not increase further when combining all features (0.80±0.10; Supplementary Figure 3). Predictions from the RF classifiers using the corrected data demonstrated a median accuracy, sensitivity and specificity of 0.70±0.10, 0.60±0.20, and 0.80±0.13, respectively.

Furthermore, a direct pairwise comparison across ROC curves revealed that the models using only the right thalamic features were generally outperformed by those using features extracted by both thalami (p's < 0.001) and, except when using only the mean signal in the model (p = 1), the left thalamic features (p's < 0.001; Table 2). Additional analyses including genotype and age as nuisance covariates yielded almost identical results (Supplementary Table 3).

Figure 4 shows the median SHAP values reflecting the impact of the different thalamic radiomic features used in the classifiers on the model output, sorted in descending rank order. Right and left mean thalamic SUVR had both a relatively high impact (compatibly with the results of the voxelwise and ROI analyses, which showed group differences in these features). However, additional radiomic features (e.g., maximum, kurtosis, entropy, energy) appeared to have comparable, or even slightly superior, impact, thus underlying that greater

information richness can be extracted from the [^{11}C]-PBR28 PET signal as compared to employing its mean value only. Similar results were observed after correcting for age and genotype (Supplementary Figure 4).

4. DISCUSSION

In spite of a large preclinical literature supporting a role for neuroinflammation in pain, so far direct evidence linking glial activation with human chronic pain has been scant. In the last few years, the use of PET radioligands targeting the 18kDa translocator protein (TSPO) has begun to fill this gap. TSPO is a five-transmembrane domain protein mostly located at the sites of contact between outer and inner mithocondrial membranes [41]. While TSPO is ubiquitous in multiple cell types throughout the body, it is widely used as a marker of inflammation because its expression levels are low in the healthy central nervous system, but become dramatically upregulated during neuroinflammatory responses, as well as highly colocalized with activated microglia/macrophages and astrocytes [41]. The colocalization between TSPO upregulation and activated microglia and astrocytes has been confirmed in multiple studies, across numerous rodent and primate models of neurodegenerative disorders, viral infection, neurotoxicity and ischemia, as well as in various human disorders [1; 9; 10; 16–18; 22; 26; 30; 36; 37; 43; 47; 48]. The association between TSPO upregulation and glial responses was also observed in a rat model of neuropathic pain [50], in which the number of TSPO-expressing astrocytes and microglia was found to increase after spinal nerve ligation (~three- and ~seven-fold compared to sham, respectively). In the same study, the proporton of TSPO-expressing neurons did not change and remained at ~5%. Because of the properties of TSPO expression, radioligands targeting this protein have the ability to image glial activation in vivo.

Using [^{11}C]-PBR28, a second-generation TSPO radioligand [12; 13; 24], we were able to demonstrate TSPO signal elevations in the brain of cLBP patients, in the medial thalamus and sensorimotor cortices [32]. In the present study we aimed to evaluate to which extent this neuroinflammatory signal is a reliable feature of cLBP. Because the number of participants investigated in the initial study was small (10 patients and 9 controls), the first aim was to replicate the original observations in independent cohorts of cLBP patients and controls. Indeed, we were able to confirm increased radioligand binding in the thalamus, using both ROI and voxel-based analyses. Notably, while similar PET signal elevations were observed in both left and right thalami in ROI analyses, the voxel-wise analyses demonstrated a left-sided dominance, which was also observed in our original publication. Since solid evidence linking pain processing or neuroinflammatory responses with left thalamic dominance has not yet been reported, this asymmetry in the thalamic neuroinflammatory signal observed in cLBP merits further exploration.

While the main observation of elevated thalamic TSPO signal has been confirmed in the replication cohort, it is also worth mentioning that we observed differences between the results of our two studies. For instance, in the current study we did not observe TSPO elevations in the primary somatosensory and motor cortices. Moreover, the local maxima of the cluster demonstrating statistically significant group differences in voxelwise analyses were not observed exactly in the same location as in the original study. One explanation may

involve differences in the clinical characteristic of the patient groups evaluated in each of the two studies. For instance, the replication cohort also included patients with only axial cLBP, whereas the previous study only included patients with some leg pain (whether with or without a radicular component). Future, larger studies should be conducted to assess whether different neuroinflammatory signatures may accompany different cLBP phenotypes. Finally, we could not replicate the association between thalamic TSPO signal and pain or serum IL-6 levels. While differences in cohort composition may explain this negative result, we also cannot discount the possibility that the (weak, albeit statistically significant) correlations observed in the original study might have been spurious, emphasizing the importance of replication studies.

Notwithstanding these differences, the replication of the overall thalamic TSPO elevations provides confidence in the notion that neuroinflammation might indeed occur in human chronic pain, as predicted by preclinical studies [27]. In this context, we recently reported [^{11}C]-PBR28 elevations in other pain conditions, including in the brain of patients with fibromyalgia [3] or migraine with aura [5], and in the spinal cord and dorsal root ganglia / nerve roots of patients with lumbar radiculopathy [2]. Of note, while TSPO signal elevations can be observed in different pain disorders, they might present condition-specific spatial distributions. For instance, thalamic neuroinflammation appears to be a hallmark of cLBP, whereas in our studies patients with migraine or fibromyalgia demonstrated limited extensive cortical increases in TSPO signal, with seemingly little or no involvement of subcortical structures. This observation raises the possibility that region specific TSPO imaging might be used to contribute to the development of objective biomarkers for different pain conditions.

As a preliminary step towards this goal, a second aim of the study was to assess to which extent the mean thalamic TSPO signal was the only informative feature that could be extracted from the combined datasets. To this end, we trained a series of RF-based machine learning algorithms based on PET imaging features, and evaluated their ability to distinguish cLBP patients from healthy controls in a predictive context. We then computed SHAP values to explore the contribution of radiomic features to the model output. These analyses revealed that, while the mean is an important feature, other features demonstrate similar level of importance, including the maximum, kurtosis, entropy, as well as others. These observations demonstrate that thalamic TSPO data have a richness of information that goes beyond the simple mean, and therefore our prior analyses have only begun to unveil.

This approach led to a median (across folds) accuracy, sensitivity and specificity of 0.70±0.10, 0.80±0.20, and 0.80±0.20, respectively, as well as an AUC of 0.88±0.07 (indicating good, bordering on excellent, performance [23]). Similar values were obtained when correcting for age and genotype (i.e., 0.70±0.10, 0.60±0.20, 0.80±0.13 and 0.80±0.10, respectively). It is important to note that, while our main focus in the classification task was on quantifying and comparing the relative importance of radiomics features, the latter were extracted from a region which showed significant group-level differences in the replication dataset. Given that our study does not include an independent test set, the absolute prediction performances are not completely unbiased and should be interpreted with caution. Still, they can serve as an initial steppingstone for future glial PET-based studies with independent data

splits and a broader focus on explainable AI [8]. Such studies are necessary to provide a complete picture on the performance and decision-making process embedded in machine learning models, in order to potentially advance imaging features such as the ones employed in our study towards biomarker status.

The focus on a single region, the thalamus, and a small number of features extracted, was adopted in order to prevent overfitting [44], particularly given the relatively small sample in our study. Nonetheless, while mean [^{11}C]-PBR28 SUVR showed a signifiicant contribution to the model, additional features appear to contribute in a similar extent, underscoring the added value of computing additional image-derived features to be used in classification analyses. These findings suggest that not only the mean uptake in PET signal, but also the characteristics of its distribution (e.g. signal heterogeneity vs homogeneity) can be harnessed to distinguish patients with cLBP from HC. Thus, it seems probable that the exploration of a full set of radiomic features (including from additional brain regions) in a future, larger study would improve the performance of the classification.

There are several limitations to our study. First, the sample size used in this study, although in line with many other PET studies, was relatively small (which may explain the relatively high dispersion in our classification performance metrics). Still, the ability to confirm the presence of a statistically significant elevation in thalamic TSPO signal in a relatively small replication cohort is encouraging, and a testament to the reliability of the observation. Furthermore, it is important to stress that the development of a “pain biomarker” was beyond the scope of this study, and that classification algorithms were employed with the sole purpose of evaluating the amount of information provided by different imaging features. A study with a much larger sample size, as well as independent test dataset, would be necessary to assess to which extent TSPO imaging metrics could be used to objectively discriminate cLBP patients from healthy controls. Moreover, a direct comparison with data from patients suffering from other pain disorders will be necessary in order to evaluate the specificity of our findings to cLBP. Another caveat to be considered is that thalamic TSPO signal did not correlate with symptom severity in the replication cohort. Therefore, the clinical significance of this signal in cLBP, and other conditions as well, awaits further investigations.

In summary, our observations suggest that thalamic neuroinflammatory signal is a reproducible and discriminating feature for chronic low back pain, and contains a richness of information that the simple mean might fail to completely capture. This work further supports a role for glial activation in human chronic low back pain, and the exploration of immunomodulatory therapies as a therapeutic approach for chronic pain.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

We thank Atreyi Saha (Athinoula A. Martinos Center for Biomedical Imaging) for helping with data management. This work was partially supported by National Institutes of Health (NIH), National Institute of Neurological

Disorders and Stroke (NINDS) 1R21NS087472-01A1 (MLL), 1R01NS095937-01A1 (MLL), 1R01NS094306-01A1 (MLL), National Cancer Institute (NCI) U01CA154601 (JKC), U24CA180927 (JKC), U24CA180918 (JKC), F30CA239407 (KC), National Institute of Biomedical Imaging and Bioengineering (NIBIB) 5T32EB1680 (KC); National Center for Complementary and Integrative Health (NCCIH), NIH (P01-AT009965), National Science Foundation (NSF) NSF1622542 (JKC), and Department of Defense (DoD) W81XWH-14-1-0543 (MLL). This research was carried out in whole or in part at the Athinoula A. Martinos Center for Biomedical Imaging at the Massachusetts General Hospital, using resources provided by the Center for Functional Neuroimaging Technologies, P41EB015896, a P41 Biotechnology Resource Grant supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB), National Institutes of Health.

REFERENCES

[1]. Abourbeh G, Theze B, Maroy R, Dubois A, Brulon V, Fontyn Y, Dolle F, Tavitian B, Boisgard R. Imaging microglial/macrophage activation in spinal cords of experimental autoimmune encephalomyelitis rats by positron emission tomography using the mitochondrial 18 kDa translocator protein radioligand [(1)(8)F]DPA-714. J Neurosci 2012;32(17):5728–5736. [PubMed: 22539835]

[2]. Albrecht DS, Ahmed SU, Kettner NW, Borra RJH, Cohen-Adad J, Deng H, Houle TT, Opalacz A, Roth SA, Melo MFV, Chen L, Mao J, Hooker JM, Loggia ML, Zhang Y. Neuroinflammation of the spinal cord and nerve roots in chronic radicular pain patients. Pain 2018;159(5):968–977. [PubMed: 29419657]

[3]. Albrecht DS, Forsberg A, Sandström A, Bergan C, Kadetoff D, Protsenko E, Lampa J, Lee YC, Höglund CO, Catana C, Cervenka S, Akeju O, Lekander M, Cohen G, Halldin C, Taylor N, Kim M, Hooker JM, Edwards RR, Napadow V, Kosek E, Loggia ML. Brain glial activation in fibromyalgia - A multi-site positron emission tomography investigation. Brain Behav Immun 2019;75:72–83. [PubMed: 30223011]

[4]. Albrecht DS, Kim M, Akeju O, Torrado-Carvajal A, Edwards RR, Zhang Y, Bergan C, Protsenko E, Kucyi A, Wasan A, Hooker J, Napadow V, Loggia ML. The neuroinflammatory component of negative affect in patients with chronic pain. Mol Psychiatr 2019;Accepted.

[5]. Albrecht DS, Mainero C, Ichijo E, Ward N, Granziera C, Zürcher NR, Akeju O, Bonnier G, Price J, Hooker JM, Napadow V, Loggia ML, Hadjikhani N. Imaging of neuroinflammation in migraine with aura: A [11C] PBR28 PET/MRI study. Neurology 2019;92(17):e2038–e2050. [PubMed: 30918090]

[6]. Albrecht DS, Normandin MD, Shcherbinin S, Wooten DW, Schwarz AJ, Zürcher NR, Barth VN, Guehl NJ, Akeju O, Atassi N, Veronese M, Turkheimer F, Hooker JM, Loggia ML. Pseudoreference Regions for Glial Imaging with 11C-PBR28: Investigation in 2 Clinical Cohorts. J Nucl Med 2018;59(1):107–114. [PubMed: 28818984]

[7]. Alshikho MJ, Zurcher NR, Loggia ML, Cernasov P, Reynolds B, Pijanowski O, Chonde DB, Izquierdo Garcia D, Mainero C, Catana C, Chan J, Babu S, Paganoni S, Hooker JM, Atassi N. Integrated magnetic resonance imaging and [(11) C]-PBR28 positron emission tomographic imaging in amyotrophic lateral sclerosis. Annals of neurology 2018;83(6):1186–1197. [PubMed: 29740862]

[8]. Arrieta A, Diaz-Rodriguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 2020;58:82–115.

[9]. Banati RB, Cagnin A, Brooks DJ, Gunn RN, Myers R, Jones T, Birch R, Anand P. Long-term trans-synaptic glial responses in the human thalamus after peripheral nerve injury. Neuroreport 2001;12(16):3439–3442. [PubMed: 11733686]

[10]. Banati RB, Newcombe J, Gunn RN, Cagnin A, Turkheimer F, Heppner F, Price G, Wegner F, Giovannoni G, Miller DH, Perkin GD, Smith T, Hewson AK, Bydder G, Kreutzberg GW, Jones T, Cuzner ML, Myers R. The peripheral benzodiazepine binding site in the brain in multiple sclerosis: quantitative in vivo imaging of microglia as a measure of disease activity. Brain 2000;123 (Pt 11):2321–2337. [PubMed: 11050032]

[11]. Breiman L Random forests. Machine Learning 2001;45(1):5–32.

[12]. Briard E, Zoghbi SS, Imaizumi M, Gourley JP, Shetty HU, Hong J, Cropley V, Fujita M, Innis RB, Pike VW. Synthesis and evaluation in monkey of two sensitive 11C-labeled aryloxyanilide ligands for imaging brain peripheral benzodiazepine receptors in vivo. J Med Chem 2008;51(1):17–30. [PubMed: 18067245]

[13]. Brown AK, Fujita M, Fujimura Y, Liow JS, Stabin M, Ryu YH, Imaizumi M, Hong J, Pike VW, Innis RB. Radiation dosimetry and biodistribution in monkey and man of 11C-PBR28: a PET radioligand to image inflammation. J Nucl Med 2007;48(12):2072–2079. [PubMed: 18006619]

[14]. Calvo M, Dawes JM, Bennett DL. The role of the immune system in the generation of neuropathic pain. Lancet Neurol 2012;11(7):629–642. [PubMed: 22710756]

[15]. Catana C, Procissi D, Wu Y, Judenhofer MS, Qi J, Pichler BJ, Jacobs RE, Cherry SR. Simultaneous in vivo positron emission tomography and magnetic resonance imaging. Proc Natl Acad Sci U S A 2008;105(10):3705–3710. [PubMed: 18319342]

[16]. Chen MK, Baidoo K, Verina T, Guilarte TR. Peripheral benzodiazepine receptor imaging in CNS demyelination: functional implications of anatomical and cellular localization. Brain 2004;127(Pt 6):1379–1392. [PubMed: 15069023]

[17]. Chen MK, Guilarte TR. Imaging the peripheral benzodiazepine receptor response in central nervous system demyelination and remyelination. Toxicol Sci 2006;91(2):532–539. [PubMed: 16554315]

[18]. Cosenza-Nashat M, Zhao ML, Suh HS, Morgan J, Natividad R, Morgello S, Lee SC. Expression of the translocator protein of 18 kDa by microglia, macrophages and astrocytes based on immunohistochemical localization in abnormal human brain. Neuropathol Appl Neurobiol 2009;35(3):306–328. [PubMed: 19077109]

[19]. Criminisi A, Shotton J. Decision forests for computer vision and medical image analysis: Springer Science & Business Media, 2013.

[20]. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837–845. [PubMed: 3203132]

[21]. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278(2):563–577. [PubMed: 26579733]

[22]. Gulyas B, Makkai B, Kasa P, Gulya K, Bakota L, Varszegi S, Beliczai Z, Andersson J, Csiba L, Thiele A, Dyrks T, Suhara T, Suzuki K, Higuchi M, Halldin C. A comparative autoradiography study in post mortem whole hemisphere human brain slices taken from Alzheimer patients and age-matched controls using two radiolabelled DAA1106 analogues with high affinity to the peripheral benzodiazepine receptor (PBR) system. Neurochem Int 2009;54(1):28–36. [PubMed: 18984021]

[23]. Hampel H, Toschi N, Baldacci F, Zetterberg H, Blennow K, Kilimann I, Teipel SJ, Cavedo E, Melo Dos Santos A, Epelbaum S, Lamari F, Genthon R, Dubois B, Floris R, Garaci F, Lista S, (APMI) APMI. Alzheimer's disease biomarker-guided diagnostic workflow using the added value of six combined cerebrospinal fluid candidates: Aβ. Alzheimers Dement 2018;14(4):492–501. [PubMed: 29328927]

[24]. Imaizumi M, Briard E, Zoghbi SS, Gourley JP, Hong J, Musachio JL, Gladding R, Pike VW, Innis RB, Fujita M. Kinetic evaluation in nonhuman primates of two new PET ligands for peripheral benzodiazepine receptors in brain. Synapse 2007;61(8):595–605. [PubMed: 17455247]

[25]. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. Neuroimage 2012;62(2):782–790. [PubMed: 21979382]

[26]. Ji B, Maeda J, Sawada M, Ono M, Okauchi T, Inaji M, Zhang MR, Suzuki K, Ando K, Staufenbiel M, Trojanowski JQ, Lee VM, Higuchi M, Suhara T. Imaging of peripheral benzodiazepine receptor expression as biomarkers of detrimental versus beneficial glial responses in mouse models of Alzheimer's and other CNS pathologies. J Neurosci 2008;28(47):12255–12267. [PubMed: 19020019]

[27]. Ji RR, Berta T, Nedergaard M. Glia and pain: is chronic pain a gliopathy? Pain 2013;154 Suppl 1:S10–28. [PubMed: 23792284]

[28]. Jung C, Ichesco E, Ratai EM, Gonzalez RG, Burdo T, Loggia ML, Harris RE, Napadow V. Magnetic resonance imaging of neuroinflammation in chronic pain: a role for astrogliosis? Pain 2020.

[29]. Kreisl WC, Jenko KJ, Hines CS, Lyoo CH, Corona W, Morse CL, Zoghbi SS, Hyde T, Kleinman JE, Pike VW, McMahon FJ, Innis RB, Team BCPRP. A genetic polymorphism for translocator protein 18 kDa affects both in vitro and in vivo radioligand binding in human brain to this putative biomarker of neuroinflammation. J Cereb Blood Flow Metab 2013;33(1):53–58. [PubMed: 22968319]

[30]. Kuhlmann AC, Guilarte TR. Cellular and subcellular localization of peripheral benzodiazepine receptors after trimethyltin neurotoxicity. J Neurochem 2000;74(4):1694–1704. [PubMed: 10737628]

[31]. Le TT, Fu W, Moore JH. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. Bioinformatics 2020;36(1):250–256. [PubMed: 31165141]

[32]. Loggia ML, Chonde DB, Akeju O, Arabasz G, Catana C, Edwards RR, Hill E, Hsu S, Izquierdo-Garcia D, Ji RR, Riley M, Wasan AD, Zürcher NR, Albrecht DS, Vangel MG, Rosen BR, Napadow V, Hooker JM. Evidence for brain glial activation in chronic pain patients. Brain 2015;138(Pt 3):604–615. [PubMed: 25582579]

[33]. Lois C, Gonzalez I, Izquierdo-Garcia D, Zurcher NR, Wilkens P, Loggia ML, Hooker JM, Rosas HD. Neuroinflammation in Huntington's Disease: New Insights with (11)C-PBR28 PET/MRI. ACS chemical neuroscience 2018;9(11):2563–2571. [PubMed: 29719953]

[34]. Lundberg SM, Su-In L. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 2017. pp. 4765–4774.

[35]. Lyoo CH, Ikawa M, Liow JS, Zoghbi SS, Morse CL, Pike VW, Fujita M, Innis RB, Kreisl WC. Cerebellum Can Serve As a Pseudo-Reference Region in Alzheimer Disease to Detect Neuroinflammation Measured with PET Radioligand Binding to Translocator Protein. J Nucl Med 2015;56(5):701–706. [PubMed: 25766898]

[36]. Maeda J, Higuchi M, Inaji M, Ji B, Haneda E, Okauchi T, Zhang MR, Suzuki K, Suhara T. Phase-dependent roles of reactive microglia and astrocytes in nervous system injury as delineated by imaging of peripheral benzodiazepine receptor. Brain Res 2007;1157:100–111. [PubMed: 17540348]

[37]. Martin A, Boisgard R, Theze B, Van Camp N, Kuhnast B, Damont A, Kassiou M, Dolle F, Tavitian B. Evaluation of the PBR/TSPO radioligand [(18)F]DPA-714 in a rat model of focal cerebral ischemia. J Cereb Blood Flow Metab 2010;30(1):230–241. [PubMed: 19794397]

[38]. Melzack R The short-form McGill Pain Questionnaire. Pain 1987;30(2):191–197. [PubMed: 3670870]

[39]. Niemann K, Mennicken VR, Jeanmonod D, Morel A. The Morel stereotactic atlas of the human thalamus: atlas-to-MR registration of internally consistent canonical model. Neuroimage 2000;12(6):601–616. [PubMed: 11112393]

[40]. Owen DR, Yeo AJ, Gunn RN, Song K, Wadsworth G, Lewis A, Rhodes C, Pulford DJ, Bennacef I, Parker CA, StJean PL, Cardon LR, Mooser VE, Matthews PM, Rabiner EA, Rubio JP. An 18-kDa translocator protein (TSPO) polymorphism explains differences in binding affinity of the PET radioligand PBR28. J Cereb Blood Flow Metab 2012;32(1):1–5. [PubMed: 22008728]

[41]. Papadopoulos V, Baraldi M, Guilarte TR, Knudsen TB, Lacapère JJ, Lindemann P, Norenberg MD, Nutt D, Weizman A, Zhang MR, Gavish M. Translocator protein (18kDa): new nomenclature for the peripheral-type benzodiazepine receptor based on its structure and molecular function. Trends Pharmacol Sci 2006;27(8):402–409. [PubMed: 16822554]

[42]. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011;12:2825–2830.

[43]. Rojas S, Martin A, Arranz MJ, Pareto D, Purroy J, Verdaguer E, Llop J, Gomez V, Gispert JD, Millan O, Chamorro A, Planas AM. Imaging brain inflammation with [(11)C]PK11195 by PET and induction of the peripheral-type benzodiazepine receptor after transient focal ischemia in rats. J Cereb Blood Flow Metab 2007;27(12):1975–1986. [PubMed: 17457364]

[44]. Theodoridis S, Koutroumbas K. Pattern Recognition: Elsevier Academic Press, 2018.

[45]. Tsuda M, Shigemoto-Mogami Y, Koizumi S, Mizokoshi A, Kohsaka S, Salter MW, Inoue K. P2X4 receptors induced in spinal microglia gate tactile allodynia after nerve injury. Nature 2003;424(6950):778–783. [PubMed: 12917686]

[46]. Turkheimer FE, Rizzo G, Bloomfield PS, Howes O, Zanotti-Fregonara P, Bertoldo A, Veronese M. The methodology of TSPO imaging with positron emission tomography. Biochem Soc Trans 2015;43(4):586–592. [PubMed: 26551697]

[47]. Venneti S, Wang G, Wiley CA. Activated macrophages in HIV encephalitis and a macaque model show increased [3H](R)-PK11195 binding in a PI3-kinase-dependent manner. Neurosci Lett 2007;426(2):117–122. [PubMed: 17888571]

[48]. Vowinckel E, Reutens D, Becher B, Verge G, Evans A, Owens T, Antel JP. PK11195 binding to the peripheral benzodiazepine receptor as a marker of microglia activation in multiple sclerosis and experimental autoimmune encephalomyelitis. J Neurosci Res 1997;50(2):345–353. [PubMed: 9373043]

[49]. Watkins LR, Hutchinson MR, Ledeboer A, Wieseler-Frank J, Milligan ED, Maier SF. Norman Cousins Lecture. Glia as the “bad guys”: implications for improving clinical pain control and the clinical utility of opioids. Brain Behav Immun 2007;21(2):131–146. [PubMed: 17175134]

[50]. Wei XH, Wei X, Chen FY, Zang Y, Xin WJ, Pang RP, Chen Y, Wang J, Li YY, Shen KF, Zhou LJ, Liu XG. The upregulation of translocator protein (18 kDa) promotes recovery from neuropathic pain in rats. J Neurosci 2013;33(4):1540–1551. [PubMed: 23345228]

[51]. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. Neuroimage 2014;92:381–397. [PubMed: 24530839]

[52]. Yoder KK, Nho K, Risacher SL, Kim S, Shen L, Saykin AJ. Influence of TSPO genotype on 11C-PBR28 standardized uptake values. J Nucl Med 2013;54(8):1320–1322. [PubMed: 23785173]

[53]. Zurcher NR, Loggia ML, Lawson R, Chonde DB, Izquierdo-Garcia D, Yasek JE, Akeju O, Catana C, Rosen BR, Cudkowicz ME, Hooker JM, Atassi N. Increased in vivo glial activation in patients with amyotrophic lateral sclerosis: assessed with [(11)C]-PBR28. Neuroimage Clin 2015;7:409–414. [PubMed: 25685708]

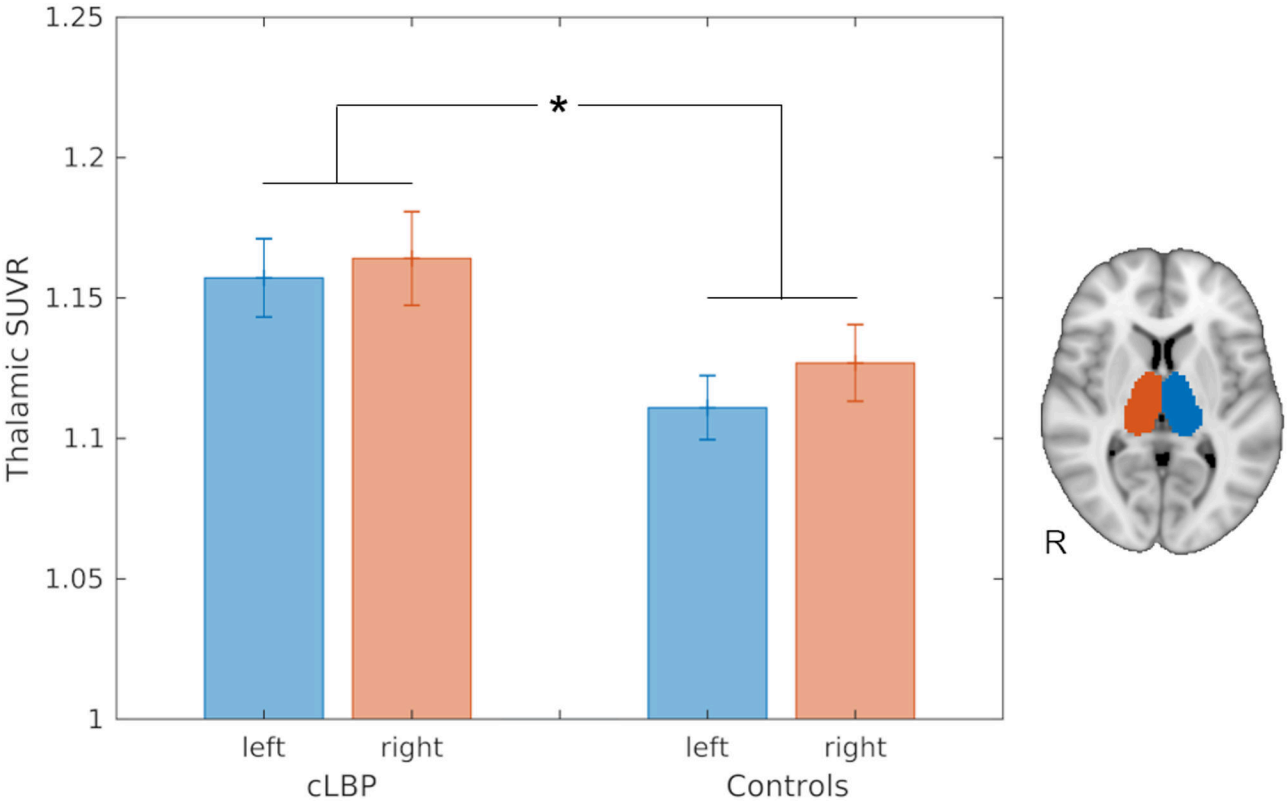

Figure 1.

Barplots displaying the mean [^{11}C]-PBR28 SUVR in the thalamic ROIs extracted from the replication cohorts (15 patients with cLBP and 17 control subjects), and corrected for genotype and age. * Group effect: $p<0.05$. Bars represent mean ± standard error of the mean (SEM).

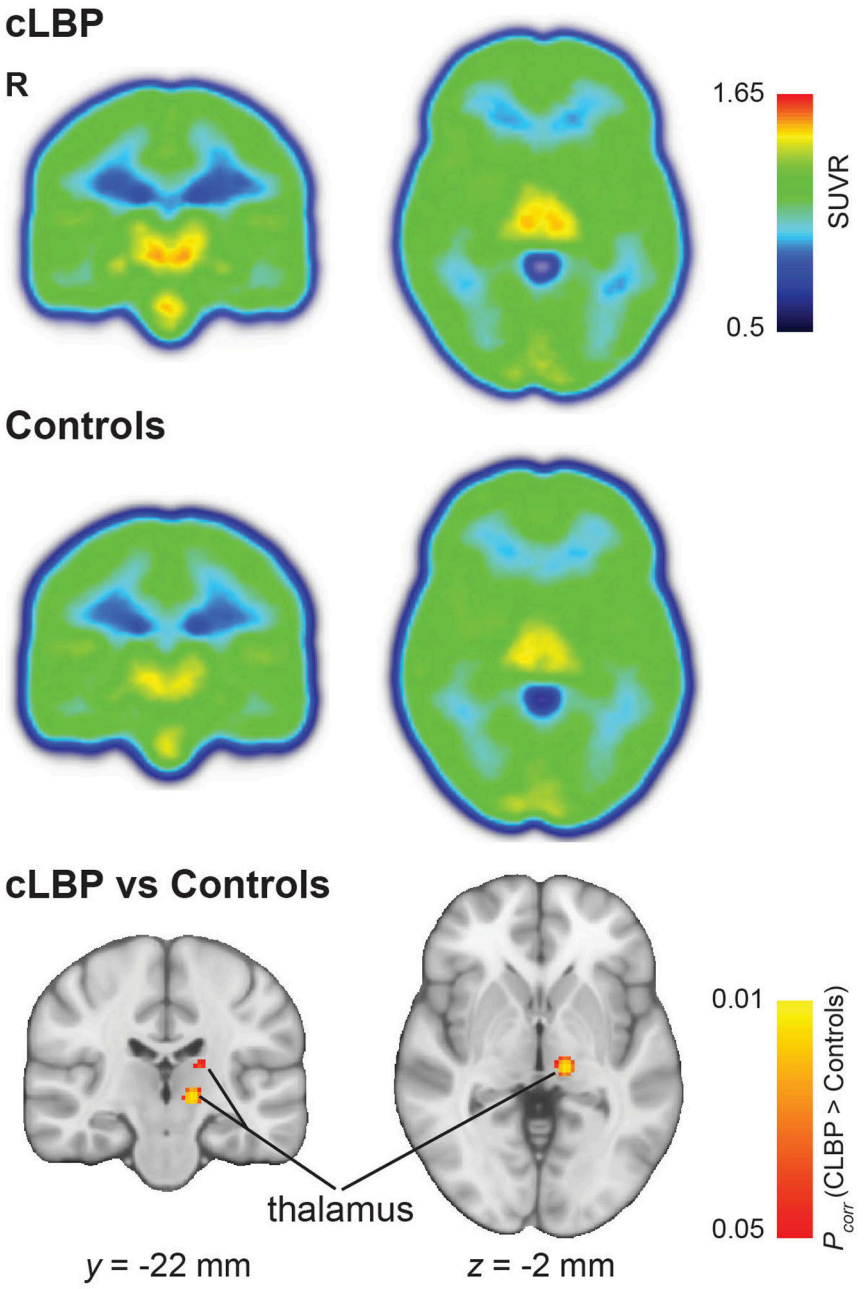

Figure 2.

Median SUVR map from patients with cLBP (upper row) and from controls (central row). Non-parametric permutation inference in the thalamus revealed significantly higher TSPO levels in patients ($p < 0.05$ corrected for multiple comparisons; bottom row).

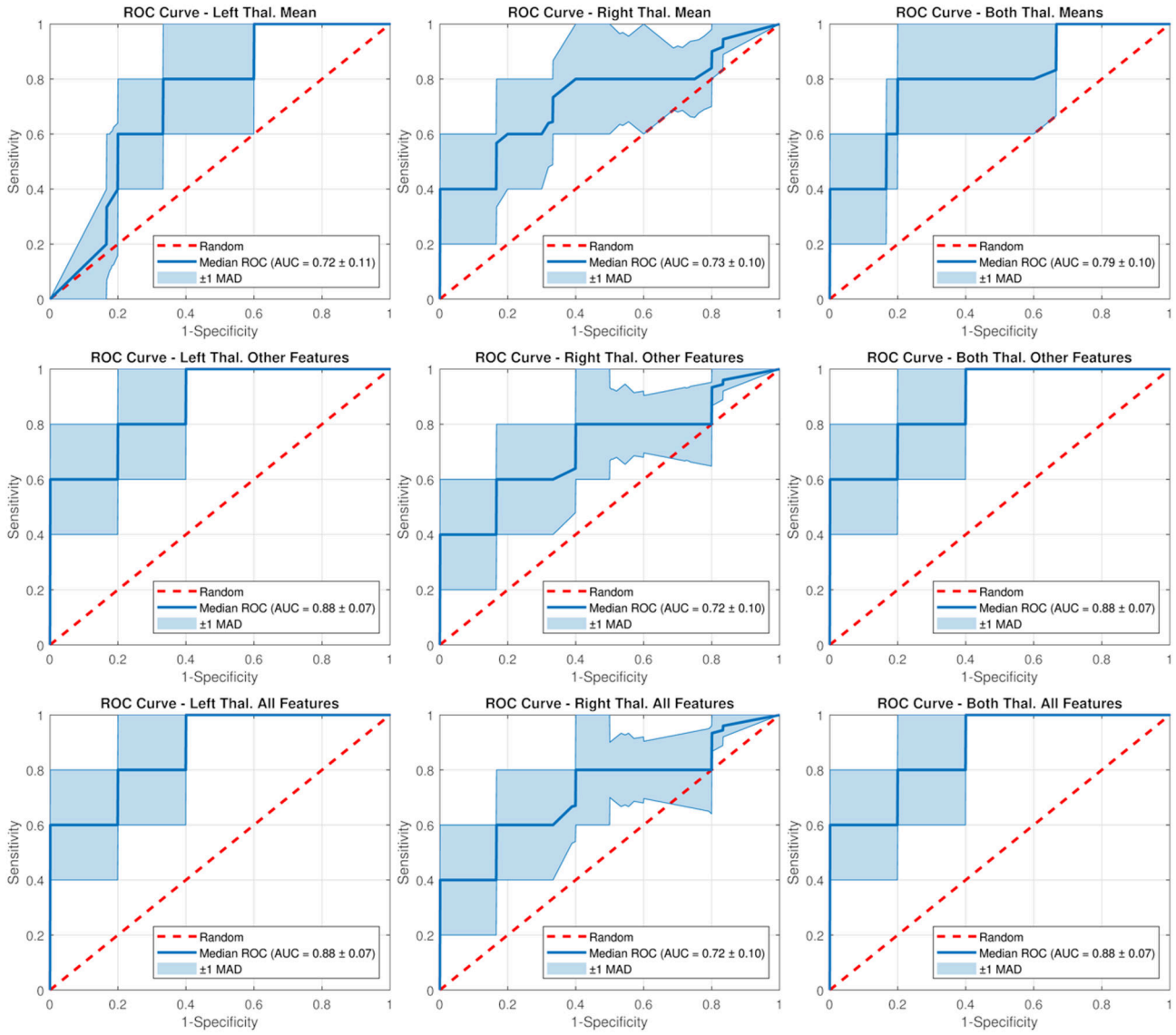

Figure 3.

Median receiver operating characteristic (ROC) curves and areas under the curve (AUCs) for the random forest classifiers selected by the genetic search based on each subset of features in the occlusion tests. MAD: Median Absolute Deviation across folds.

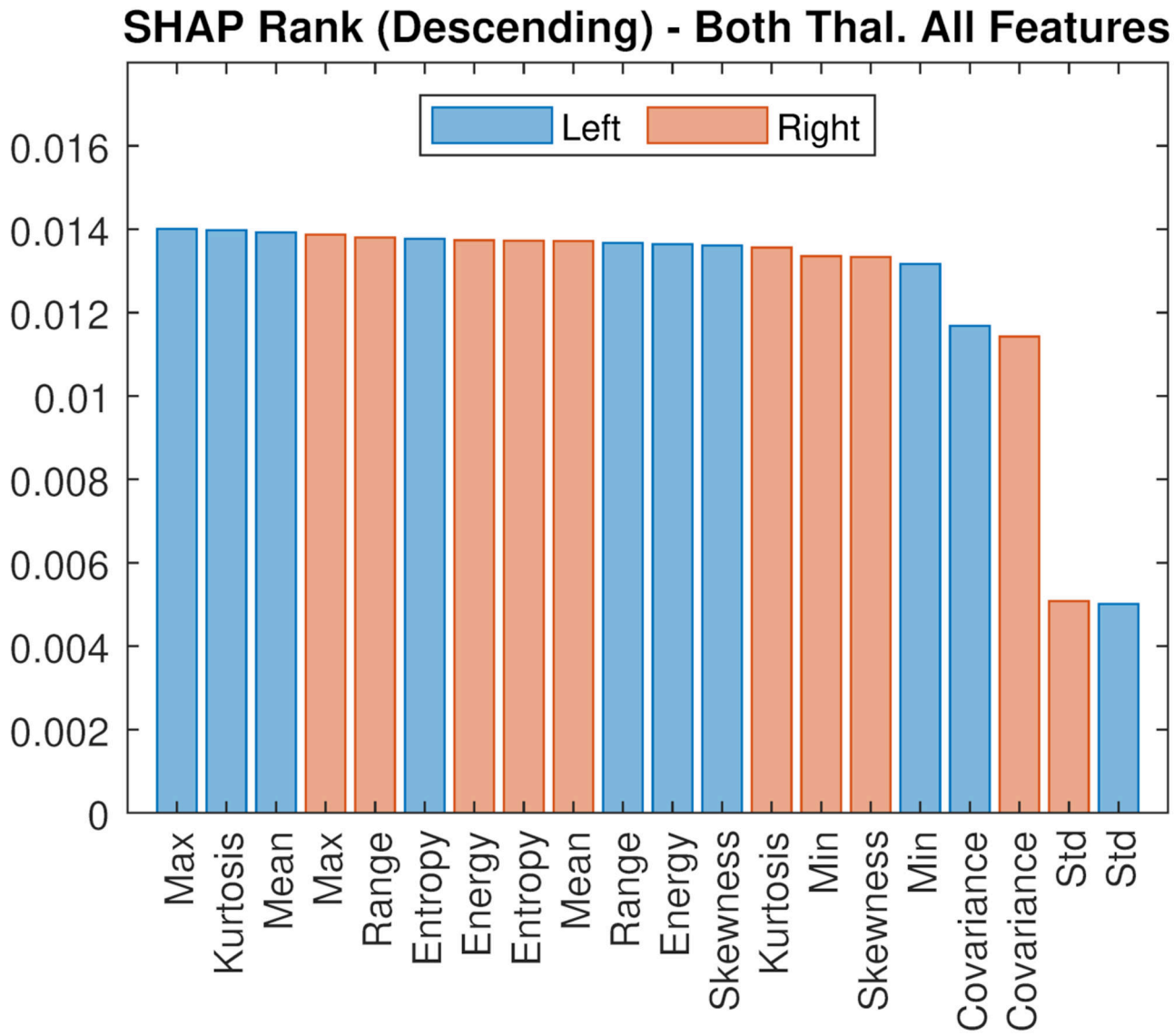

Figure 4.
SHAP values reflecting features impact of the different thalamic radiomic features used in the classifiers on the model output, sorted in descending rank order. Bars represent the median.

Table 1.

Direct pairwise comparison across ROC curves to explore the richness and ability of different features of the neuroinflammatory signal, in alternative to or in combination with the mean, to discriminate cLBP patients from controls. The table shows the resulting *p*-values, corrected for multiple comparisons using Bonferroni correction, obtained when comparing AUC differences in ROC curves from different classifiers using a paired DeLong test.

	Mean vs Others	Mean vs All	Others vs All
Left	< 0.001	< 0.001	1
Right	< 0.001	< 0.001	1
Both	< 0.001	< 0.001	0.07

Table 2.

Direct pairwise comparison across ROC curves to explore the richness and ability of left thalami, in alternative to or in combination with the right thalami, to discriminate cLBP patients from controls. The table shows the resulting *p*-values, corrected for multiple comparisons using Bonferroni correction, obtained when comparing AUC differences in ROC curves from different classifiers using a paired DeLong test.

	Left > Right	**Left < Both**	**Right < Both**
Mean	1	< 0.001	< 0.001
Others	< 0.001	1	< 0.001
All	< 0.001	1	< 0.001