Research Paper

PAIN 164 (2023) 2516–2527

PAIN®

OPEN

Brain morphology predicts individual sensitivity to pain: a multicenter machine learning approach

Raviteja Kotikalapudi[a], Balint Kincses[a,b], Matthias Zunhammer[b], Frederik Schlitt[b], Livia Asan[b], Tobias Schmidt-Wilcke[c,d], Zsigmond T. Kincses[e,f], Ulrike Bingel[b], Tamas Spisak[a,*]

Abstract

Sensitivity to pain shows a remarkable interindividual variance that has been reported to both forecast and accompany various clinical pain conditions. Although pain thresholds have been reported to be associated to brain morphology, it is still unclear how well these findings replicate in independent data and whether they are powerful enough to provide reliable pain sensitivity predictions on the individual level. In this study, we constructed a predictive model of pain sensitivity (as measured with pain thresholds) using structural magnetic resonance imaging–based cortical thickness data from a multicentre data set (3 centres and 131 healthy participants). Cross-validated estimates revealed a statistically significant and clinically relevant predictive performance (Pearson $r = 0.36$, $P < 0.0002$, $R^2 = 0.13$). The predictions were found to be specific to physical pain thresholds and not biased towards potential confounding effects (eg, anxiety, stress, depression, centre effects, and pain self-evaluation). Analysis of model coefficients suggests that the most robust cortical thickness predictors of pain sensitivity are the right rostral anterior cingulate gyrus, left parahippocampal gyrus, and left temporal pole. Cortical thickness in these regions was negatively correlated to pain sensitivity. Our results can be considered as a proof-of-concept for the capacity of brain morphology to predict pain sensitivity, paving the way towards future multimodal brain-based biomarkers of pain.

Keywords: Pain sensitivity, Cortical thickness, Predictive modelling, Machine Learning, Crossvalidation, Quantitative Sensory Testing, Pain thresholds, Anterior cingulate cortex

1. Introduction

Pain sensitivity is known to considerably vary across individuals.[62] This variability has been reported to both forecast and accompany various clinical pain conditions[53,58] and—besides peripheral aspects—it seems to be shaped to a large degree by individual differences in brain structure and function.[66,84,93,99]

Recent studies found that task-based[99,105] and resting-state[47,84] brain function, as measured with magnetic resonance imaging (MRI), has the potential to characterize pain on the individual level, with effect sizes that may be sufficient for aiding diagnosis and for evaluation of risk of developing pain and of analgesic efficacy.[92] Structural T1-weighted MRI is a promising additional modality for the brain-based characterization of pain, with a potentially high translational value, because of its reliability,[40,43,55] relatively shorter scanning times (only 5-6 minutes for a complete MR-structural sequence), and its wide availability in clinical routine (a structural scan is a standard prerequisite for fMRI/diffusion-weighted images). Several studies have reported structural brain correlates of pain sensitivity on the group level. Voxel-based morphometry (VBM) studies have reported that pain sensitivity is significantly correlated to the morphology of various structures, including the cingulate cortex (CC), precuneus, primary somatosensory cortex (S1), putamen, insula, and parahippocampal gyrus (PHG), among others.[22,60,61,71,103] Surface-based morphometry techniques,[24] such as cortical thickness analysis, can provide additional insights and, in contrast to VBM, a better account for gyrification.[26] For instance, Erpelding et al.[23] found a predominantly positive correlation between cortical thickness and heat pain thresholds in S1, posterior midcingulate, and orbitofrontal cortices. These and similar studies have identified many possible morphological correlates of pain sensitivity. However, the partly conflicting findings in these studies suggest that we should not be overly optimistic about the replicability of these associations. In general, such mass-univariate brain-wide association studies (BWAS) has been reported to typically provide small effect sizes and often lack reliability.[51,83] However, given the known issues of replicability and small effect sizes in such mass-univariate brain-wide association studies (BWAS), it is unclear whether such "pain—brain morphology" associations can serve with replicable, biomedically relevant effect sizes.[51,83] Issues of

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

[a] Institute for Diagnostic and Interventional Radiology and Neuroradiology, University Medicine Essen, Essen, Germany, [b] Department of Neurology, Center for Translational Neuro- and Behavioural Sciences, University Medicine Essen, Essen, Germany, [c] Institute for Clinical Neuroscience and Medical Psychology, Heinrich Heine University, Düsseldorf, Germany, [d] Neurocenter, District Hospital Mainkofen, Deggendorf, Germany, Departments of [e] Neurology and, [f] Radiology, University of Szeged, Szeged, Hungary

**Corresponding author. Address: Institute for Diagnostic and Interventional Radiology and Neuroradiology, University Medicine Essen, 45147 Essen, Germany. Tel.: +49 201/723-6379. E-mail address: tamas.spisak@uk-essen.de (T. Spisak).*

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.painjournalonline.com).

http://dx.doi.org/10.1097/j.pain.0000000000002958

replicability and small effect sizes can be overcome by multivariate predictive models that integrate small effects across brain regions into robust predictions with larger effect sizes and offer substantial promise for delivering robust and replicable brain-based measures of various behavioural and clinical states and traits, including pain.[104] Yet, we lack systematically evaluated and externally validated brain morphology-based predictive models for pain.[83,102]

We aim to develop a cortical thickness-based predictive model that is (1) specific to individual pain sensitivity and not confounded by other related but out-of-interest factors,[17,33,42,77,82] (2) generalizable for heterogenous out-of-centre data,[98,102] and (3) explainable, ie, it provides insights into the underlying mechanism.[84,99]

2. Methods

2.1. Participants

In total, n = 133 participants were recruited, and experimental procedures were performed at 3 different study sites, namely, Ruhr University Bochum, Germany (study 1, n = 39); University Medicine Essen, Germany (study 2, n = 49); and University of Szeged, Hungary (study 3, n = 45). Reimbursement was performed only for study 1 and study 2 at 20 €/h. All studies were conducted in accordance with the Declaration of Helsinki and ethically approved by the local or national committees (register Numbers: 4974-14, 18-8020-BO, 057617/2015/OTIG, and ETT TUKEB for studies 1-3, respectively). Inclusion and exclusion criteria are explained in detail elsewhere.[84] In brief, inclusion criteria comprised no chronic diseases, age ≥18 and ≤40 years (target age = 25 years), right-handed, nonsmoking participants, and a balanced sex distribution. Moreover, participants were excluded in the presence of acute or chronic neurological, endocrine, or psychiatric disorders; acute infections; the use of psychotropic or analgesic-based substances; wounds; scars; or skin irritations, that could potentially affect the pain stimuli-based experimental setup. Magnetic resonance imaging scans underwent a quality check procedure to exclude images with incomplete whole brain coverage or motion artefacts (supplementary Figure 1, available at http://links.lww.com/PAIN/B858). Overall, n = 2 MRI scans with an incomplete frontal brain acquisition were excluded from final analysis (n = 131).

2.2. Pain sensitivity measure assessment based on pain thresholds

Heat (HPT), cold (CPT), and mechanical pain thresholds (MPT) of the participants were assessed following the established quantitative sensory testing (QST) protocol.[68] Although warmth (WDT) and cold detection thresholds (CDT) were examined in study 1 and study 2, in addition, mechanical detection thresholds (MDT) were also collected for study 3. Thermal thresholds were obtained using advanced thermal stimulators (ATS thermodes) on a skin surface of 30 × 30 mm at the volar forearm. For thermal stimulators, MSA thermal stimulator (Somedic, Horby, Sweden) was used in study 1, and pathway thermal stimulators (Medoc Ltd, Ramat Yishai, Israel) were used in study 2 and 3. Increasing and decreasing thermal thresholds were applied to the skin, and the baseline temperature was kept at 32°C. Using a button press, participants indicated heat and cold pain onsets. Instead of 3 (original protocol), 6 stimuli repetitions were conducted for all thermal thresholds to address within-subject variability. To determine mean WDT, CDT, HPT, and CPT, we calculated the arithmetic means for each threshold. Therefore, only data of the measurements 2 to 6 were used. The first measurement of each WDT, CDT, HPT, and CPT assessment was defined as a test measurement and, therefore, excluded from the analysis. The MPT was measured using a set of 7 pin-prick mechanical stimulators with fixed stimulus intensities (flat contact area of 0.2 mm diameter) that exerted forces of 8, 16, 32, 64, 128, 256, and 512 mN. The stimulators were applied at a rate of 2 seconds on and 2 seconds off in an ascending order until the first percept of sharpness was reached. The final threshold was calculated as the log-transformed geometric mean of 5 series of ascending and descending stimuli. Mechanical pain thresholds and MDTs were determined using a staircase method. Five increasing and 5 decreasing trains of pinprick (MRC Systems, Heidelberg, Germany) stimuli were applied to the palmar left forearm in an alternating fashion, whereas the participant was instructed to categorize the stimuli as painful or not painful. Mechanical detection threshold was assessed analogously with von Frey filament stimulations. Mechanical pain threshold and MDT were computed as the log-transformed geometric mean force determined in 5 ascending and descending staircase-thresholding runs. Finally, a QST score was calculated as a composite score composed of HPT, CPT, and MPT. All 3 measures were z-transformed within each centre. Heat pain threshold and MPT were inverted (× −1; the algebraic signs "−" were adjusted so that it reflects the overall pain sensitivity of the participant in the final composite score, ie, QST score). After this, we calculated the arithmetic mean of the 3 pain thresholds for each participant. Quantitative sensory testing scores that were ≥ 2.5 standard deviations from the mean were defined as outliers and should, therefore, be excluded. None of the QST scores were identified as outliers.

2.3. Additional measures

We acquired additional measures wherever possible at the 3 study centres to evaluate potential confounders. Additional measures comprised age, body mass index (BMI), date of the first day of the last menses (females), level of education (primary, secondary, and university), and the following self-report questionnaires: Pain Sensitivity Questionnaire: PSQ,[70] Pain Catastrophizing Scale: PCS,[86] Pittsburgh Sleep Quality Index: PSQI,[10] Perceived Stress Questionnaire: PSQ20,[49] State-Trait Anxiety Inventory: STAI,[81] and short German version of the Depression Scale (Centre for Epidemiological Studies-Depression Scale): ADS-K.[48] Blood pressure was measured both before the MRI and QST procedure. In study 1, so-called T_{50} values, ie, temperatures referring to a pain intensity of 50 on a 0 to 100 Visual Analogue Scale (VAS; anchors: 0 = no pain and 100 = intolerable pain), were additionally measured. These measures along with WDT, CDT, and MDT measurements were used to analyse the confounding effects on the model that was based on predicting the composite pain sensitivity score (QST, for testing model specificity). For study specific sample sizes, please refer to supplementary table 1, available at http://links.lww.com/PAIN/B858.

2.4. Magnetic resonance imaging data

Our study uses 3D T_1-weighted (structural) images in all study centres, performed on 3T scanners with an isotropic voxel size of 1 mm^3. To avoid the "standardization fallacy"—ie, to preserve a certain level of data heterogeneity so that the findings could generalize towards future image acquisitions (Voelkl et al. 2021)—no further standardization of imaging sequences has been performed across sites. Site-specific scanning parameters are

listed in **Table 1**. The 3 centres used different scanners, namely, Philips Achieva (study 1), Siemens Magnetom Skyra (study 2), and GE Discovery 750 w MR (study 3). Structural image processing of T_1-weighted images was conducted by using the FreeSurfer software, version 6.0 (available at https://surfer.nmr.mgh.harvard.edu/). The underlying working framework for image processing using FreeSurfer is covered extensively elsewhere.[12,14,25–30] We decided to restrict our analyses to gray matter (GM) thickness and exclude other measures of morphology (eg, GM volume, area, and density or morphological connectivity). We selected gray matter thickness as a measure because of its reliability and the quantitative nature of this measure,[38] as well as its potential independence from factors such as the brain surface area and head size in comparison with GM volume, where factor correction procedures are not well standardized and could induce unnecessary statistical artifacts.[4,73] Each processing steps involved in obtaining cortical thickness using recon-all measures are listed with explanations at https://surfer.nmr.mgh.harvard.edu/fswiki/recon-all. In brief, native space images were transformed to a standard space using a Talairach transformation.[87] Further image processing steps included correction for motion artifacts,[67] skull stripping,[74] removal of the cerebellum and brain stem, correction of nonuniform intensities,[78] segmentation of subcortical, white matter (WM), and deep gray matter (GM) regions, that helps in estimating GM-WM junction, tessellation of GM-WM boundary, smoothing of tessellated surface, automated topology correction,[75] and generating a model for pial surface (GM-cerebrospinal fluid junction).[14,25] Cortical thickness was measured as the average of distance to the nearest point on the pial surface from each vertex of the tessellated WM boundary and from that point back to the nearest point on the WM boundary. Cortical thickness was measured as the average of distance to the nearest point on the pial surface from each vertex of the tessellated WM boundary and from that point back to the nearest point on the WM boundary. Cortical thickness estimates were averaged within the regional parcellations of the Desikan–Killiany atlas,[16] as (1) it is a widely accepted atlas used in many multicenter collaborative studies involving cortical thickness measures[46,101] and (2) it offers a favourable balance between the number of features (68) and our sample size (131), which can help to mitigate issues of overfitting or inadequate model fitting.[37] The overall procedure for obtaining cortical thicknesses has been previously validated against histological and manual measurements and its reliability has been established across different scanning parameters such as MRI sequences and scanning machines[41,69,72].We considered the cortical thickness measures from the FreeSurfer analysis, which resulted in a total of 68 regional thickness measures (34 per hemisphere, measured in millilitres).

2.5. Feature harmonization

The feature set comprised 68 features (cortical thickness in 68 regions of the Desikan–Killiany atlas[16]) for 131 subjects. Regional cortical thickness measures were scaled by normalizing with the mean cortical thickness at the per participant. As cortical thickness measures were previously reported to be highly specific to scanning sites (Fortin et al.[31]), the feature set was harmonized to address the site effects. The harmonization process was performed using ComBat, a batch-effect correction tool frequently used in genomics and multicenter neuroimaging studies.[46] ComBat uses a Bayesian approach to enhance the reliability of estimated parameters in smaller samples and has been shown to be more effective than using simple feature residuals, phenotype-adjusted residuals, or adding site as a regressor to the predictive model.[11,17,31,82] ComBat was configured to mitigate site effects, while still preserving the biologically relevant effects of age and sex. To avoid feature leakage,[52] harmonization has been incorporated into the machine learning cross-validation framework (see next paragraph, Machine learning pipeline). Leakage was prevented by using 2 functions: ComBat fit and ComBat transform. The fit function can be used on the (outer) training set to fit a ComBat model, whereas the transform function can be used to simply apply the already fitted ComBat transformation on the hold-out feature set. As a result, the transformed training set is guaranteed not to learn any information from its corresponding test set, thus avoiding data leakage, which could potentially bias the model predictions.

2.6. Machine learning pipeline

The machine learning (ML) framework is depicted in **Figure 1**. First, the entire sample (feature set: 131 × 68 and corresponding target set = QST scores, 132 × 1) was split into train (feature set + target) and its corresponding test (only feature set) sets using a balanced 10-fold cross validation (CV, using GroupKFold: 1-fold corresponds to 1 iteration), where each fold would hold approximately the same amount of data from all 3 samples. This CV framework would serve as the outer loop for our ML pipeline. In this outer loop, the train data set is fit + transformed using ComBat, and the test set is transformed with the already fitted ComBat model, as described above. In the inner CV-loop, the train set from the outer loop is further split into a train subset and a validation subset in a 10-fold CV fashion. Subsequently, in each fold, the transformed train subset with its corresponding target values (ie, QST-based composite pain sensitivity scores) was used to fit a linear regression model with least absolute shrinkage and selection operator (LASSO),[90] using the scikit-learn,[63] a python-based package (https://scikit-learn.org/stable/). LASSO shrinks the features of less importance and, potentially, eliminates them (nullifying coefficient/weight of the feature) to reduce model complexity and prevent

Table 1

Magnetic resonance imaging acquisition information.

	Study 1	Study 2	Study 3
Scanner	Philips Achieva X	Siemens Magnetom Skyra	GE Discovery MR750w
Head coil	32-channel	32-channel	20-channel
Sequence	MP-RAGE	MP-RAGE	IR-FSPGR
FOV	256 × 256 × 220 mm^3	256 × 256 × 192 mm^3	256 × 256 × 172 mm^3
TR	8500 ms	2300 ms	5.3 ms
TE	3.9 ms	2.07 ms	2.1 ms

Study specific MRI acquisition information is presented including scanner protocols.
FOV, field of view; IF-FSPGR, inversion recovery fast spoiled gradient echo; MP-RAGE, magnetization prepared rapid gradient echo; MRI, magnetic resonance imaging; TE, echo time; TR, repetition time.

overfitting. LASSO's tendency to fully eliminate unimportant features can improve the interpretability of the resulting model but will inevitably result in a shrinkage of the magnitude of the predictions, as compared with the magnitude of the target variable. Shrinkage in LASSO is implemented by an L1-penalty equal to the absolute value of the magnitude of feature coefficient and can be adjusted through tuning a single hyperparameter: α (learning rate). At different hyperparameter values for α ([0.0001, 0.001, 0.01, 0.1, 1, 10, 1000, and 10000]), the LASSO model was fit to the inner train set and evaluated on the validation subset for mean squared error (MSE) when predicting the composite pain sensitivity scores. The best LASSO model estimator in each inner loop was applied to create predictions for the actual outer test set (final predictions) in each of the 10 outer CV loops. As our initial model was confounded by estimated total intracranial volume (eTIV) ($P = 0.006$, see 2.7. confounder analysis for methodological details), we controlled for eTIV post hoc at the level of machine learning predictions, as proposed by (Dinga et al. 2020). In details, eTIV was regressed out from the output of the LASSO model in a cv-consistent fashion, ie, the regression coefficients were fit on the train set predictions (in each cv iteration), and in the test set, the unconfounded predictions were calculated based on these coefficients. Overall, any potential cofounding effects with the final predictions were regressed out.[17] Finally, we evaluated feature importance for each "best estimator model," corresponding to the 10 outer CV iterations and considered features, which have a nonzero coefficient to be robust predictors for pain sensitivity. Pearson r was calculated for all associations, and its P-value was determined using nonparametric permutation testing with 10,000 permutations.

2.7. Confounder analyses

To evaluate potential confounding bias in the model predictions for centre-based, demographic-based, and psychological-based additional measures, we used the "mlconfound" package, which is available from https://mlconfound.readthedocs.io.[82] We performed the partial confounder test with the null hypothesis of "no confounder bias," which is tested by probing the conditional independence of the predicted pain sensitivity on each additional measure, given the observed pain sensitivity. The test is distinguished from alternative approaches by its robustness to nonnormally and nonlinearly dependent predictions. Therefore, it is applicable without investigating the conditional distribution of the predictions on the target and the confounder for normality and linearity. A P-value < 0.05 implies confounding bias (ie, the predictions are driven partially by the confounder).

2.8. Leave-one-study-out

Along with the main analysis, we also assessed out-of-centre generalizability by means of a leave-one-study-out crossvalidation, ie, predicting each study by fitting the above-described machine learning model on the remaining 2 studies. Pain sensitivity in the left-out centre was predicted with the model that performed best on data from the remaining 2 centres (with hyperparameter α optimized in a nested cross-validation loop). That means, study 1 + 2 (1 + 3 and 2 + 3) were used to estimate pain sensitivity in a completely unseen study 3 (2 and 1). Parameters, such as α values in the inner CV data splitting scheme (GroupKFold(10)), remained the same as mentioned in the previous section (machine learning pipeline). It should be noted, that, since ComBat can only consider batch IDs that it has already seen during training, the final predictions were performed twice, ie, the left-out study was considered by ComBat to be from an identical source as either the first or the second from the training studies (and mean predictions were considered).

2.9. Correlation with the Human Connectome Project 1200 data

To further characterize the most important predictors, we analyzed already processed FreeSurfer data from the Human Connectome Project (HCP; www.humanconnectome.org, WU-Minn HCP 1200 Subjects Data Release, release of the data: March 1, 2017). Although the HCP1200 does not involve measures of pain sensitivity, it contains 2 pain-related scores, namely, the pain intensity and pain interference surveys (both self-reported questionnaires by the participant). The data for the 2 questionnaires is collected by the pain intensity survey for age >18 years (The NIH Toolbox Sensation Measures, www.nihtoolbox.org). The survey consists of 1 self-report item asking about the participant's level of pain in the past 7 days. Participants were asked to rate their pain on a scale of 0 (no pain) to 10 (worst imaginable pain). Associations were tested with Pearson correlation. We tested only those regions that emerged as the nonzero-weighted features (a priori) in our LASSO-driven predictive model. The correlations' one-sided P-value was obtained using nonparametric permutation testing (n = 10,000). Like our data analysis, the cortical thickness values were normalized with mean thickness and area measures with eTIV. Before performing the correlations, cortical measures were adjusted for age, sex, TIV, and batch effects using ComBat.

3. Results

Our predictive model has been developed in a sample of n = 131 healthy participants, recruited at 3 different scanning centres. We trained a multivariate model on normalized mean cortical thickness values in 68 unilateral, anatomically defined brain regions (based on the Desikan–Killiany brain atlas[16]) to predict the QST-based[68] composite pain threshold score based on threshold sensitivity, as defined in previous studies.[84,106] Model characterization involved brain morphological data from a total of n = 1226 participants (including data from the Human Connectome Project[94]) and focused on 3 aspects: (1) estimation of the unbiased effect size of the multicentre predictive model with a balanced, nested cross-validation strategy; (2) testing out-of-centre generalizability on each of the 3 independent imaging centres; and (3) evaluation of the neuroscientific validity of the model and the specificity of its predictions to physical pain thresholds, by investigating model bias towards a comprehensive set of "validator variables," including demographics, psychometrics, sensory thresholds, neurotransmitter levels and subjective, and self-evaluative pain phenotypes (which are known to significantly diverge from physical pain thresholds[13,19]). Finally, by analysing the predictive coefficients, we identified 3 key regions that drive the morphology-based pain sensitivity predictions: right rostral anterior cingulate cortex (rACC), the left PHG, and the left temporal pole (TP).

3.1. Model predictions

The multicentre model predicted pain sensitivity in unseen participants with a medium effect size (Pearson correlation coefficient $r = 0.36$, P-value = 0.0002, $R^2 = 0.13$). The predicted values also correlated significantly with the QST scores from each individual site, ie, study 1 ($r = 0.42$, $P = 0.01$, $R^2 = 0.18$ see Supplementary Figure 2a–c, available at http://links.lww.com/PAIN/B858), study 2 ($r = 0.32$, $P = 0.01$, $R^2 = 0.10$), and study 3 ($r = 0.35$, $P = 0.01$, $R^2 = 0.12$). Multicenter model predictions for

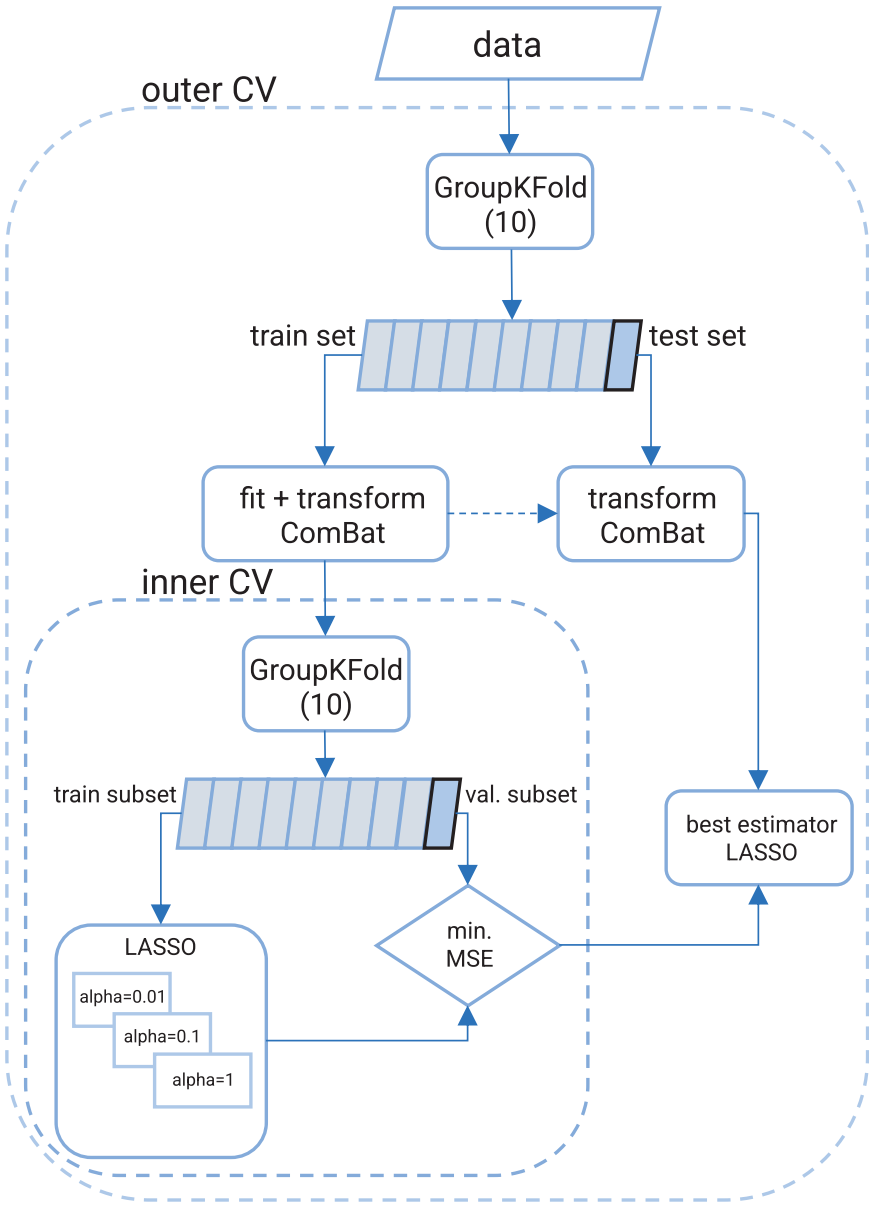

Figure 1. Flow chart of the machine learning pipeline. The flowchart shows the splitting of multicentre data into train and test sets and subsequent processes. In the outer crossvalidation (CV), train sets consist of participant's cortical thickness matrix and QST-scores, which are used for fitting LASSO (least absolute shrinkage and selection operator) models. In the test sets, the participants' cortical thickness measures are fed into the already fitted LASSO model (optimized in the inner CV loop) to provide crossvalidated predictions for the pain sensitivity scores. The nested cross-validation scheme and the use of the fit-transform scheme for data harmonization (ComBat) avoid feature leakage and provides unbiased estimates of the predictive performance. MSE, mean squared error; QST, quantitative sensory testing.

pain sensitivity are presented in **Figures 2A and B**, supplementary figure-table 3, available at http://links.lww.com/PAIN/B858.

3.2. Model properties

The best model was found with an $\alpha = 0.01$ (regularization constant hyperparameter) consequently across all 10 iterations of the outer loop crossvalidation in the multicentre analysis. The applied machine learning model (LASSO) eliminates features of less importance by assigning a feature weight (coefficient) = 0. Hence, we interpret the features with nonzero weights (coefficients) as the most prominent regions involved in predicting pain sensitivity. There were 6 regions that had nonzero coefficients in at least one of the cross-validation iterations and 3 of them emerged in all iterations (**Table 2**). These 3 regions also displayed the highest predictive importance in the model (mean predictive coefficient, averaged over crossvalidation iterations). Based on absolute weights (coefficients of the feature), the right rACC provided the highest contribution to the model's predictions, which was followed by left parahippocampal gyrus (PHG) and left TP. All 3 features were negatively associated with the pain sensitivity scores (**Figs. 2C–E**), ie, right rACC ($r = -0.38$, $P = 0.0001$), left PHG ($r = -0.24$, $P = 0.003$), and TP ($r = -0.16$, $P = 0.03$), indicating that thinner cortices were associated with higher pain sensitivity. Three more regions, namely, the right PHG ($r = -0.22$, $P = 0.01$), right frontal pole ($r = 0.12$, $P = 0.09$), and left entorhinal cortex ($r = -0.13$, $P = 0.07$), have also emerged with nonzero coefficients in at least one CV iteration, but their appearance was not consistent as they emerged only in 1 (right PHG), 5 (right frontal pole), and 2 (left entorhinal cortex) CV iterations, respectively.

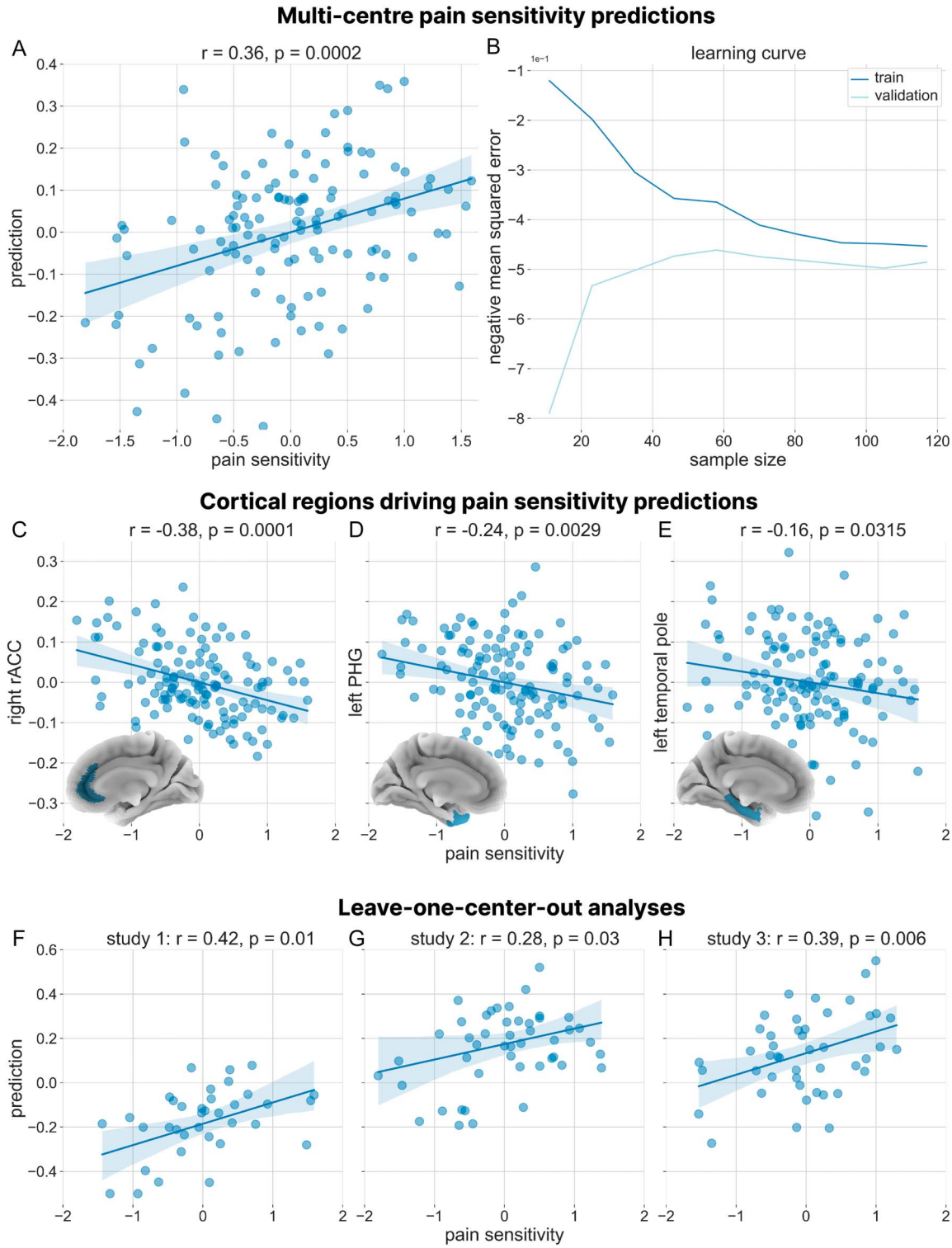

Figure 2. Gray matter cortical thickness predicts pain sensitivity. Scatter plots and regression lines (with 95% confidence intervals) depicting the correlation between pain sensitivity and the machine learning model's predictions are shown in this figure. (A) The pooled multicentre data set (balanced nested 10-fold crossvalidation), (B) learning curve for the best model for the train and validation sets suggests that increasing training sample size would result in marginal benefits for model performance, (C–E) the main cortical regions driving predictions, ie, right rostral anterior cingulate cortex, left parahippocampal gyrus, and left temporal pole, respectively, (F–H) results from the leave-one-center/study-out analyses. r = Pearson correlation coefficient, P = P-value tested for 10,000 permutations. PHG, parahippocampal gyrus; rACC, rostral anterior cingulate cortex.

3.3. Leave-one-center-out validation

Leave-one-study-out analyses revealed a significant model performance on out-of-centre data. (**Figs. 2F–H**) and Supplementary figure 1, available at http://links.lww.com/PAIN/B858 and, for the coefficients of predictive regions, **Table 3**) Predictions across all 3 studies were significant ($P < 0.05$), with effect sizes comparable with

Table 2

Cross-validation feature information.

	Right rACC	Left PHG	Left TP	Right FP	Right PHG	Left ENT
Mean weight	−1.67	−0.53	−0.21	0.11	−0.02	−0.01
STD	0.24	0.18	0.10	0.11	—	0.001
Frequency (/total CV)	10/10	10/10	10/10	5/10	1/10	2/10
Feature rank	1	2	3	—	—	—

The predictive features are remarkably robust across crossvalidation folds. Mean and standard deviation of the model coefficients across all outer CV iterations are shown for all features that were given a nonzero coefficient at least in one crossvalidation fold. The frequency of occurrence in the model across all CV iterations reveals the right rACC, the left PHG, and the left TP as the most robust predictors of pain sensitivity. The hyperparameter remained constant across all 10 crossvalidations at $\alpha = 0.01$.

CV, crossvalidation; ENT, entorhinal cortex; FP, frontal pole; PHG, parahippocampal gyrus; rACC, rostral anterior cingulate cortex; STD, standard deviation; TP, temporal pole.

that of the multicentre prediction. Common predictor across all studies was the right rACC. The left PHG was an additional common predictor for study 1 and 3 (right PHG in case of study 2), whereas the left TP was observed to be an additional common predictor for study 2 and 3. Other predictors included the right frontal pole (FP) (study 1) and the left entorhinal cortex (study 3).

3.4. Confounder analyses

The cofounder analysis of the final model revealed no significant (all P-value > 0.05) confounding bias for any of the investigated demographic, physiological, and psychological variables (**Table 4**).

3.5. Correlation with the human connectome project 1200 data

We checked for associations of the morphological measures (cortical thickness and area) of the 3 most predictive regions (rACC, TP, and PHC) and the self-reported pain intensity and inference in the HCP1200 data set (Supplementary table 2, available at http://links.lww.com/PAIN/B858). We did not find any associations between pain and the rACC. A weak but significant negative correlation was observed between the thickness of the TP and the pain intensity scores ($r = -0.06$, $P = 0.02$, $R^2 = 0.0036$) on both sides. For the pain interference T score, there was a weak but significant negative correlation with the area of PHG ($r = -0.06$, $P = 0.02$, $R^2 = 0.0036$) that was found bilaterally.

4. Discussion

We have developed a multicentre predictive model for pain sensitivity assessed by pain threshold measurements based on structural brain morphology. The proposed model can be considered as a proof-of-concept for the capacity of brain morphology to yield robust and specific individual-level predictions for pain sensitivity, posing brain structure as a promising component of future multimodal brain-based pain biomarkers.[47,91] Our results broaden our knowledge about the brain structural correlates of individual differences and alterations in pain sensitivity by identifying 3 key regions that drive the morphology-based pain sensitivity predictions: the right rACC, the left parahippocampal gyrus (PHG), and the left temporal pole (TP).

Robust, clinically relevant predictive performance is a key requirement for clinically useful predictive models. In our study, grey matter cortical thickness explained 13% of the variance in pain sensitivity ($R^2 = 0.13$ for the single, cross-validated model based on all centres and $R^2 = 0.08$-0.17 in the leave-one-centre-out cross-validation). Both according to Cohen rules of thumb, and in comparison with typical effect sizes in multivariate brain-wide association studies (eg, mean $r \approx 0.1$ when predicting cognition from cortical thickness),[51] this is a substantial predictive effect size that may hold clinical relevance[18] and demonstrates that brain morphology has the potential to be a valuable modality for objectively characterizing pain based on brain-based features. Our model can serve as a baseline for future studies that, with the aid of highly optimized models (eg, advanced morphological features assessed in a larger number of brain regions), will likely yield even higher predictive effect sizes.

The applied leakage-free nested cross-validation framework with a minimal, conventional feature preprocessing approach mitigates effect size inflation that could stem from methodological flexibility choices ("vibration effects," ie, ratio of largest versus smallest effect on the same associations explored through different model parameters).[39,95] Moreover, the model training procedure proved to be highly stable for hyperparameters (regularization) and model coefficients (predictive weights) when generalizing to the heterogeneous left-out data set (different scanners, sequences, research staff, etc). Although this robustness may justify cautious optimism regarding the generalizability of the proposed model to external data sets, a comprehensive evaluation of generalizability requires independent external validation studies with large samples. To facilitate future external validation studies, we provide a containerized version of our model, to compute pain sensitivity predictions from BIDS-formatted data sets in a single step (https://github.com/pni-lab/ctp-signature).

Clinically relevant predictive performance and generalizability to new data are necessary but not sufficient criteria for brain-based biomarker candidates.[82,102] Specificity to pain and the absence of bias towards potential confounders are also crucial factors that determine the clinical and translational utility of brain-

Table 3

Out-of-centre generalization results.

Left-out-study-ID	Coefficients
Study 1	Right rACC = −1.72, left PHG = −0.49, right FP = 0.14
Study 2	Right rACC = −1.39, left TP = −0.89, right PHG = −0.87
Study 3	Right rACC = −1.86, left PHG = −0.77, left ENT = −0.29, left TP = −0.06

The regularization factor remained constant in all validations at $\alpha = 0.01$. Study 1 (2 and 3, respectively) was always a completely new data set to the model that was calculated using study 2 and 3 (1 and 3 and 1 and 2, respectively). TIV was regressed out from final predictions within the cross-validation framework. Age and gender were adjusted along with the batch (study centre) in the ComBat model before predictions.

ENT, entorhinal cortex; PHG, parahippocampal gyrus; rACC, rostral anterior cingulate cortex; TIV, total intracranial volume; TP, temporal pole; FP, frontal pole.

Table 4

Results of the confounder analysis for predictions with additional measures.

No.	Validator variable	Correlation with pain sensitivity (*r*)	Correlation with model predictions (*r*)	Significance of confounder bias (*P*)
1	BP at MRI systole	−0.11	−0.07	0.53
2	BP at MRI diastole	0.01	0.13	0.20
3	BP at QST systole	−0.1	−0.03	0.82
4	BP at QST diastole	−0.05	0.04	0.72
5	MRI-QST time diff.	0.06	0.07	0.48
6	Sex	−0.09	−0.02	0.88
7	Day of menses	−0.09	0.05	0.72
8	Age	0.03	−0.07	0.42
9	BMI	0.01	0.04	0.71
10	Education	−0.05	−0.13	0.26
11	Alcohol per unit	0.15	0.21	0.25
12	Alcohol per week	−0.05	−0.05	0.62
13	PCS: catastrophizing	0.09	0.03	0.72
14	PCS: Rumination	0.14	0.01	0.93
15	T50	−0.45	0.12	0.71
16	GLX	0.36	0.37	0.10
17	GABA	0.24	0.11	0.58
18	Anxiety state	−0.01	0.03	0.75
19	Anxiety trait	0.11	−0.02	0.81
20	Pain questionnaire	0.16	0.03	0.74
21	Depression	−0.11	−0.01	0.91
22	Stress	0.17	0.06	0.64
23	Sleep quality	−0.04	0.1	0.42
24	Cold detection thresh.	0	0.07	0.50
25	Warm detection thresh.	0	0.08	0.44
26	Mechanical detection thresh.	0.05	0.02	0.32
27	Study center	−0.01	0.04	0.65
28	Total intracranial vol.	−0.02	−0.02	0.842

Confounder analysis was performed with the partial confounder test.[82] A *P*-value < 0.05 suggests that confounder bias exists and that the predictions are not exclusively driven by the cortical thickness but also by the validator variable.

BMI, body mass index; BP, blood pressure; GABA, gamma aminobutyric acid; GLX, glutamate + glutamine; MRI, magnetic resonance imaging; *P*, *P*-value after 10,000 permutations; PCS, Pain Catastrophizing Scale; QST, quantitative sensory testing; *r*, Pearson correlation coefficient.

based predictive models.[82] Our confounder analysis showed that the predictions of the proposed model are not biased by variables-of-no-interest, such as centre effects (shown to be problematic for cortical thickness[31]), demographics, intracranial volume, blood pressure, menstrual cycle, alcohol consumption, or sleep quality. The predictions were not primarily driven by sensory detection thresholds (as measured by QST) suggesting that the model is specific to pain (as opposed to being a marker of general sensory sensitivity).

Clinically relevant predictive performance of the proposed model, its potential generalizability, as well as its high neuroscientific validity and specificity highlights that, next to functional brain signatures of pain,[84,99] brain morphology should also be considered as a promising modality for the objective brain-based characterization of pain. The reported predictive performance, together with the relatively short (approx. 5-6 minutes), reliable[73] and widely applicable and available data acquisition protocol, and the lightweight data analysis requirements render brain morphology not only as an additional modality of future multimodal pain biomarkers[91]—complementary to functional brain signatures of pain[84,99]—but also as a highly accessible, standalone tool for pain research.

Next to serving as a proof-of-concept for the utility of morphology in the construction of brain-based predictive models of pain, the present work also broadens our knowledge about the structural brain correlates of individual pain sensitivity differences. Despite the robustness of the model's predictions, it is important to note that the LASSO-assigned feature weights may be subject to uncertainty. Thus, it is crucial to interpret the model predictors within the context of their biological plausibility and not for their inferential guarantee as infallible indicators of pain sensitivity.[89] In all 3 regions that were robustly identified as valuable predictors (rACC, PHG, and TP), we found that thicker cortex was associated with lower pain sensitivity. Cortical thickness has been found to be inversely correlated with laminar differentiation,[100] myelination,[59] glial cell involvement,[97] and—in a region

dependent manner—neuronal density,[44] altogether suggesting a close coupling between cortical structure and functional demand. To this end, it is appealing to hypothesize that the observed predictive capacity of cortical thickness in these regions originates from their differential functional involvement in pain processing and the resulting structural plasticity. Consistently with this notion, various studies have shown a significant association between cortical thickness and pain in healthy participants[23] as well as in pain disorders.[15,21,50]

The predictive regions identified in this study fit well into this framework because the pain-related function of these regions is widely acknowledged. Identifying rACC thickness as the most important morphological predictor of pain sensitivity is in line with the collective evidence for the involvement of rACC in modulating nociceptive processes.[1,5–7,20,32,45,80,85,88,99] A thicker rACC might be a consequence of a higher functional involvement in nociception-related neural processes that are involved in attenuating pain. For example, Bingel et al.[9] showed that repetitive pain stimulation over a number of days did not only significantly decrease individual pain ratings but also increased the functional involvement of the rACC. Furthermore, in line with our finding, a thicker rACC cortex has been linked to lower pain sensitivity levels in long-term meditation practitioners when compared with a control group.[35] The lack of associations between rACC thickness and self-reported pain intensity or interference in the HCP data is tempting to be interpreted in the light of the previously reported robust ($N = 505$) divergence between QST-based pain thresholds and self-reported pain sensitivity,[19,36,54] with the latter being apparently more closely related to anxiety. This may suggest that the observed pain-related thickness changes in the rACC are specific to behaviour in the presence of acute painful experiences (as measured by QST) and do not generalize to the subjective beliefs about one's own sensitivity to (intangible) pain, as measured by self-reports.

Although chronic pain has been also consistently linked to lower grey matter volume in the rACC,[79] it remains unclear whether structural features of chronic conditions are based on similar histological correlates as in heightened acute pain sensitivity and if a thin rACC might even predispose for pain chronicity. Translational animal studies could help to elucidate the cellular mechanisms for these macroscopic changes.[2]

The PHG and TP were also found to be robust predictors of pain sensitivity, although with a lower predictive coefficient and weaker unimodal association with pain sensitivity, as compared with rACC ($r = -0.24$ and $r = -0.16$, respectively). Several studies have described the involvement of PHG and hippocampal networks in nociceptive processes,[8,96] and functional connectivity of these areas was posed as an important modulator of pain quality experiences, possibly mediated by anticipatory anxiety and associative learning.[64] Several structural findings underpin the notion that differences in PHG function may manifest in regional morphology. For instance, Mutso et al.[57] reported a reduction of hippocampal volume in chronic pain patients and proposed synaptic plasticity and neurogenesis as possible mechanisms, based on a neuropathic rodent pain model. Moreover, in the aforementioned study, Grant et al. reported a negative association between cortical thickness and pain sensitivity in long-term mediation practitioners not only in the rACC but, interestingly, also in the PHG.[35] Neumann et al.[60] recently also reported increased grey matter volume in PHG with lower pain sensitivity in centre 1 of the present multicentre analysis.

The predictive capacity of the thickness of the TP may also be traced back to its pain-related function. Several studies reported TP activation after noxious stimuli.[3,34,56,76] Furthermore, the TP has been brought into relation with the interaction of pain and working memory, changing the way stimuli are later remembered by acting directly on memory encoding.[34] The putative role of both the hippocampal formation and the TP in the formation of pain-related memories is underpinned by our results showing that morphological measures of these regions—but not the rACC—showed a weak but significant negative correlation with self-reported pain intensity and inference ratings in the HCP data set. As self-reported pain scores are known to be associated with both anxiety[19] and the interaction between pain and working memory (changing the way how painful experience is remembered),[34] these results support the previously reported role of hippocampal and temporal formations in the exacerbation of pain by anxiety.[65]

Altogether, we provide a robust predictive modelling of individual differences in perceiving pain, which broadens our understating of the morphological markers of pain sensitivity. The identified predictive model may have potential implications for translational pain research and novel analgesic treatment strategies in precision medicine.

Conflict of interest statement

The authors have no conflicts of interest to declare.

Acknowledgements

The research was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): TRR 289 Treatment Expectation—project number 422744262 and SFB 874 "Integration and Representation of Sensory Processes," project A8 (Tobias Schmidt-Wilcke). Data processing was executed on the high-performance computing (HPC) cluster of the Institute of Artificial Intelligence in Medicine (IKIM), University Medicine Essen and University of Duisburg-Essen, Germany.

Supplemental digital content

Supplemental digital content associated with this article can be found online at http://links.lww.com/PAIN/B858.

Article history:
Received 19 August 2022
Received in revised form 18 February 2023
Accepted 23 March 2023
Available online 6 June 2023

References

[1] Apkarian AV, Bushnell MC, Treede R-D, Zubieta J-K. Human brain mechanisms of pain perception and regulation in health and disease. Eur J Pain 2005;9:463–84.
[2] Asan L, Falfán-Melgoza C, Beretta CA, Sack M, Zheng L, Weber-Fahr W, Kuner T, Knabbe J. Cellular correlates of gray matter volume changes in magnetic resonance morphometry identified by two-photon microscopy. Scientific Rep 2021;11:1–20.
[3] Atlas LY, Lindquist MA, Bolger N, Wager TD. Brain mediators of the effects of noxious heat on pain. PAIN 2014;155:1632–48.
[4] Barnes J, Ridgway GR, Bartlett J, Henley SM, Lehmann M, Hobbs N, Clarkson MJ, MacManus DG, Ourselin S, Fox NC. Head size, age and gender adjustment in MRI studies: a necessary nuisance? Neuroimage 2010;53:1244–55.
[5] Benarroch EE. What is the role of the cingulate cortex in pain? Neurology 2020;95:729–32.
[6] Bingel U, Herken W, Teutsch S, May A. Habituation to painful stimulation involves the antinociceptive system-a 1-year follow-up of 10 participants. PAIN 2008;140:393–4.

[7] Bingel U, Lorenz J, Schoell E, Weiller C, Büchel C. Mechanisms of placebo analgesia: rACC recruitment of a subcortical antinociceptive network. PAIN 2006;120:8–15.
[8] Bingel U, Quante M, Knab R, Bromm B, Weiller C, Büchel C. Subcortical structures involved in pain processing: evidence from single-trial fMRI. PAIN 2002;99:313–21.
[9] Bingel U, Schoell E, Herken W, Büchel C, May A. Habituation to painful stimulation involves the antinociceptive system. PAIN 2007;131:21–30.
[10] Buysse DJ, Reynolds CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28:193–213.
[11] Chyzhyk D, Varoquaux G, Milham M, Thirion B. How to remove or control confounds in predictive models, with applications to brain biomarkers. Gigascience 2022;11:giac014.
[12] Collins DL, Neelin P, Peters TM, Evans AC. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J Comput Assist Tomogr 1994;18:192–205.
[13] Coronado RA, George SZ. The Central Sensitization Inventory and Pain Sensitivity Questionnaire: an exploration of construct validity and associations with widespread pain sensitivity among individuals with shoulder pain. Musculoskelet Sci Pract 2018;36:61–7.
[14] Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 1999;9:179–94.
[15] DaSilva AF, Granziera C, Snyder J, Hadjikhani N. Thickening in the somatosensory cortex of patients with migraine. Neurology 2007;69: 1990–5.
[16] Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006;31: 968–80.
[17] Dinga R, Schmaal L, Penninx BW, Veltman DJ, Marquand AF. Controlling for effects of confounding variables on machine learning predictions. BioRxiv 2020. doi: 10.1101/2020.08.17.255034.
[18] Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, Haythornthwaite JA, Jensen MP, Kerns RD, Ader DN, Brandenburg N, Burke LB, Cella D, Chandler J, Cowan P, Dimitrova R, Dionne R, Hertz S, Jadad AR, Katz NP, Kehlet H, Kramer LD, Manning DC, McCormick C, McDermott MP, McQuay HJ, Patel S, Porter L, Quessy S, Rappaport BA, Rauschkolb C, Revicki DA, Rothman M, Schmader KE, Stacey BR, Stauffer JW, von Stein T, White RE, Witter J, Zavisic S. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105–21.
[19] Edwards RR, Fillingim RB. Self-reported pain sensitivity: lack of correlation with pain threshold and tolerance. Eur J Pain 2007;11: 594–8.
[20] Eippert F, Bingel U, Schoell ED, Yacubian J, Klinger R, Lorenz J, Büchel C. Activation of the opioidergic descending pain control system underlies placebo analgesia. Neuron 2009;63:533–43.
[21] Ellerbrock I, Engel AK, May A. Microstructural and network abnormalities in headache. Curr Opin Neurol 2013;26:353–9.
[22] Emerson NM, Zeidan F, Lobanov OV, Hadsel MS, Martucci KT, Quevedo AS, Starr CJ, Nahman-Averbuch H, Weissman-Fogel I, Granovsky Y, Yarnitsky D, Coghill RC. Pain sensitivity is inversely related to regional grey matter density in the brain. PAIN 2014;155:566–73.
[23] Erpelding N, Moayedi M, Davis KD. Cortical thickness correlates of pain and temperature sensitivity. PAIN 2012;153:1602–9.
[24] Fischl B. FreeSurfer. Neuroimage 2012;62:774–81.
[25] Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci 2000;97: 11050–5.
[26] Fischl B, Liu A, Dale AM. Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans Med Imaging 2001;20:70–80.
[27] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Van Der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002;33: 341–55.
[28] Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. Neuroimage 1999;9:195–207.
[29] Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp 1999;8:272–84.
[30] Fischl B, Van Der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D. Automatically parcellating the human cerebral cortex. Cereb Cortex 2004;14:11–22.
[31] Fortin J-P, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, Adams P, Cooper C, Fava M, McGrath PJ, McInnis M, Phillips ML, Trivedi MH, Weissman MM, Shinohara RT. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage 2018; 167:104–20.
[32] Fuchs PN, Peng YB, Boyette-Davis JA, Uhelski ML. The anterior cingulate cortex and pain processing. Front Integr Neurosci 2014;8:35.
[33] Giusti EM, Lacerenza M, Manzoni GM, Castelnuovo G. Psychological and psychosocial predictors of chronic postsurgical pain: a systematic review and meta-analysis. PAIN 2021;162:10–30.
[34] Godinho F, Magnin M, Frot M, Perchet C, Garcia-Larrea L. Emotional modulation of pain: is it the sensation or what we recall? J Neurosci 2006;26:11454–61.
[35] Grant JA, Courtemanche J, Duerden EG, Duncan GH, Rainville P. Cortical thickness and pain sensitivity in zen meditators. Emotion 2010; 10:43–53.
[36] Grundström H, Larsson B, Arendt-Nielsen L, Gerdle B, Kjølhede P. Associations between pain thresholds for heat, cold and pressure, and Pain Sensitivity Questionnaire scores in healthy women and in women with persistent pelvic pain. Eur J Pain 2019;23:1631–9.
[37] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Machine Learn Res 2003;3:1157–82.
[38] Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S, Busa E, Pacheco J, Albert M, Killiany R, Maguire P, Rosas D, Makris N, Dale A, Dickerson B, Fischl B. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. Neuroimage 2006;32:180–94.
[39] Ioannidis JP. Why most discovered true associations are inflated. Epidemiology 2008;19:640–8.
[40] Iscan Z, Jin TB, Kendrick A, Szeglin B, Lu H, Trivedi M, Fava M, McGrath PJ, Weissman M, Kurian BT, Adams P, Weyandt S, Toups M, Carmody T, McInnis M, Cusin C, Cooper C, Oquendo MA, Parsey RV, DeLorenzo C. Test–retest reliability of freesurfer measurements within and between sites: effects of visual approval process. Hum Brain Mapp 2015;36: 3472–85.
[41] Jovicich J, Czanner S, Han X, Salat D, van der Kouwe A, Quinn B, Pacheco J, Albert M, Killiany R, Blacker D. MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. Neuroimage 2009;46:177–92.
[42] Klauenberg S, Maier C, Assion H-J, Hoffmann A, Krumova EK, Magerl W, Scherens A, Treede R-D, Juckel G. Depression and changed pain perception: hints for a central disinhibition mechanism. PAIN 2008;140: 332–43.
[43] Knussmann GN, Anderson JS, Prigge MB, Dean DC III, Lange N, Bigler ED, Alexander AL, Lainhart JE, Zielinski BA, King JB. Test-retest reliability of FreeSurfer-derived volume, area and cortical thickness from MPRAGE and MP2RAGE brain MRI images. Neuroimage Rep 2022;2: 100086.
[44] la Fougère C, Grant S, Kostikov A, Schirrmacher R, Gravel P, Schipper HM, Reader A, Evans A, Thiel A. Where in-vivo imaging meets cytoarchitectonics: the relationship between cortical thickness and neuronal density measured with high-resolution [18F] flumazenil-PET. Neuroimage 2011;56:951–60.
[45] Lamm C, Decety J, Singer T. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. Neuroimage 2011;54:2492–502.
[46] Larivière S, Rodríguez-Cruces R, Royer J, Caligiuri ME, Gambardella A, Concha L, Keller SS, Cendes F, Yasuda C, Bonilha L, Gleichgerrcht E, Focke NK, Domin M, von Podewills F, Langner S, Rummel C, Wiest R, Martin P, Kotikalapudi R, O'Brien TJ, Sinclair B, Vivash L, Desmond PM, Alhusaini S, Doherty CP, Cavalleri GL, Delanty N, Kälviäinen R, Jackson GD, Kowalczyk M, Mascalchi M, Semmelroch M, Thomas RH, Soltanian-Zadeh H, Davoodi-Bojd E, Zhang J, Lenge M, Guerrini R, Bartolini E, Hamandi K, Foley S, Weber B, Depondt C, Absil J, Carr SJA, Abela E, Richardson MP, Devinsky O, Severino M, Striano P, Tortora D, Hatton SN, Vos SB, Duncan JS, Whelan CD, Thompson PM, Sisodiya SM, Bernasconi A, Labate A, McDonald CR, Bernasconi N, Bernhardt BC. Network-based atrophy modeling in the common epilepsies: a worldwide ENIGMA study. Sci Adv 2020;6:eabc6457.
[47] Lee J-J, Kim HJ, Čeko M, Park B-y, Lee SA, Park H, Roy M, Kim S-G, Wager TD, Woo C-W. A neuroimaging biomarker for sustained experimental and clinical pain. Nat Med 2021;27:174–82.
[48] Lehr D, Hillert A, Schmitz E, Sosnowsky N. Screening depressiver störungen mittels allgemeiner depressions-skala (ADS-K) und state-trait depressions scales (STDS-T) eine vergleichende evaluation von Cut-off-werten. Diagnostica 2008;54:61–70.

[49] Levenstein S, Prantera C, Varvo V, Scribano ML, Berto E, Luzi C, Andreoli A. Development of the Perceived Stress Questionnaire: a new tool for psychosomatic research. J Psychosomatic Res 1993;37:19–32.
[50] Maleki N, Becerra L, Brawn J, Bigal M, Burstein R, Borsook D. Concurrent functional and structural cortical alterations in migraine. Cephalalgia 2012;32:607–20.
[51] Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, Donohue MR, Foran W, Miller RL, Hendrickson TJ, Malone SM, Kandala S, Feczko E, Miranda-Dominguez O, Graham AM, Earl EA, Perrone AJ, Cordova M, Doyle O, Moore LA, Conan GM, Uriarte J, Snider K, Lynch BJ, Wilgenbusch JC, Pengo T, Tam A, Chen J, Newbold DJ, Zheng A, Seider NA, Van AN, Metoki A, Chauvin RJ, Laumann TO, Greene DJ, Petersen SE, Garavan H, Thompson WK, Nichols TE, Yeo BTT, Barch DM, Luna B, Fair DA, Dosenbach NUF. Reproducible brain-wide association studies require thousands of individuals. Nature 2022;603:654–60.
[52] Mateos-Pérez JM, Dadar M, Lacalle-Aurioles M, Iturria-Medina Y, Zeighami Y, Evans AC. Structural neuroimaging as clinical predictor: a review of machine learning applications. Neuroimage Clin 2018;20:506–22.
[53] Meints SM, Mawla I, Napadow V, Kong J, Gerber J, Chan S-T, Wasan AD, Kaptchuk TJ, McDonnell C, Carriere J, Rosen B, Gollub RL, Edwards RR. The relationship between catastrophizing and altered pain sensitivity in patients with chronic low back pain. PAIN 2019;160:833–43.
[54] Meiselles D, Aviram J, Suzan E, Pud D, Eisenberg E. Does self-perception of sensitivity to pain correlate with actual sensitivity to experimental pain? J Pain Res 2017;10:2657–63.
[55] Melzer TR, Keenan RJ, Leeper GJ, Kingston-Smith S, Felton SA, Green SK, Henderson KJ, Palmer NJ, Shoorangiz R, Almuqbel MM, Myall DJ. Test-retest reliability and sample size estimates after MRI scanner relocation. Neuroimage 2020;211:116608.
[56] Moulton E, Becerra L, Maleki N, Pendse G, Tully S, Hargreaves R, Burstein R, Borsook D. Painful heat reveals hyperexcitability of the temporal pole in interictal and ictal migraine states. Cereb Cortex 2011;21:435–48.
[57] Mutso AA, Radzicki D, Baliki MN, Huang L, Banisadr G, Centeno MV, Radulovic J, Martina M, Miller RJ, Apkarian AV. Abnormalities in hippocampal functioning with persistent pain. J Neurosci 2012;32:5747–56.
[58] Nahman-Averbuch H, Leon E, Hunter BM, Ding L, Hershey AD, Powers SW, King CD, Coghill RC. Increased pain sensitivity but normal pain modulation in adolescents with migraine. PAIN 2019;160:1019–28.
[59] Natu VS, Gomez J, Barnett M, Jeska B, Kirilina E, Jaeger C, Zhen Z, Cox S, Weiner KS, Weiskopf N, Grill-Spector K. Apparent thinning of human visual cortex during childhood is associated with myelination. Proc Natl Acad Sci 2019;116:20750–9.
[60] Neumann L, Wulms N, Witte V, Spisak T, Zunhammer M, Bingel U, Schmidt-Wilcke T. Network properties and regional brain morphology of the insular cortex correlate with individual pain thresholds. Hum Brain Mapp 2021;42:4896–908.
[61] Niddam DM, Wang S-J, Tsai S-Y. Pain sensitivity and the primary sensorimotor cortices: a multimodal neuroimaging study. PAIN 2021;162:846–55.
[62] Nielsen CS, Staud R, Price DD. Individual differences in pain sensitivity: measurement, causation, and consequences. J Pain 2009;10:231–7.
[63] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. Scikit-learn: machine learning in Python. J Machine Learn Res 2011;12:2825–30.
[64] Pinto CB, Bielefeld J, Barroso J, Yip B, Huang L, Schnitzer T, Apkarian AV. Chronic pain domains and their relationship to personality, abilities, and brain networks. PAIN 2022;164:59–71.
[65] Ploghaus A, Narain C, Beckmann CF, Clare S, Bantick S, Wise R, Matthews PM, Rawlins JNP, Tracey I. Exacerbation of pain by anxiety is associated with activity in a hippocampal network. J Neurosci 2001;21:9896–903.
[66] Reddan MC, Wager TD. Modeling pain using fMRI: from regions to biomarkers. Neurosci Bull 2018;34:208–15.
[67] Reuter M, Rosas HD, Fischl B. Highly accurate inverse consistent registration: a robust approach. Neuroimage 2010;53:1181–96.
[68] Rolke R, Baron R, Maier Ca, Tölle T, Treede DR, Beyer A, Binder A, Birbaumer N, Birklein F, Bötefür I, Braune S, Flor H, Huge V, Klug R, Landwehrmeyer GB, Magerl W, Maihöfner C, Rolko C, Schaub C, Scherens A, Sprenger T, Valet M, Wasserka B. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. PAIN 2006;123:231–43.
[69] Rosas H, Liu A, Hersch S, Glessner M, Ferrante R, Salat D, van Der Kouwe A, Jenkins B, Dale A, Fischl B. Regional and progressive thinning of the cortical ribbon in Huntington's disease. Neurology 2002;58:695–701.
[70] Ruscheweyh R, Marziniak M, Stumpenhorst F, Reinholz J, Knecht S. Pain sensitivity can be assessed by self-rating: development and validation of the Pain Sensitivity Questionnaire. PAIN 2009;146:65–74.
[71] Ruscheweyh R, Wersching H, Kugel H, Sundermann B, Teuber A. Gray matter correlates of pressure pain thresholds and self-rated pain sensitivity: a voxel-based morphometry study. PAIN 2018;159:1359–65.
[72] Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RS, Busa E, Morris JC, Dale AM, Fischl B. Thinning of the cerebral cortex in aging. Cereb Cortex 2004;14:721–30.
[73] Schwarz CG, Gunter JL, Wiste HJ, Przybelski SA, Weigand SD, Ward CP, Senjem ML, Vemuri P, Murray ME, Dickson DW, Parisi JE, Kantarci K, Weiner MW, Petersen RC, Jack CR. A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. Neuroimage Clin 2016;11:802–12.
[74] Ségonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, Fischl B. A hybrid approach to the skull stripping problem in MRI. Neuroimage 2004;22:1060–75.
[75] Ségonne F, Pacheco J, Fischl B. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEE Trans Med Imaging 2007;26:518–29.
[76] Shinozaki T, Imamura Y, Kohashi R, Dezawa K, Nakaya Y, Sato Y, Watanabe K, Morimoto Y, Shizukuishi T, Abe O, Haji T, Tabei K, Taira M. Spatial and temporal brain responses to noxious heat thermal stimuli in burning mouth syndrome. J Dental Res 2016;95:1138–46.
[77] Sivertsen B, Lallukka T, Petrie KJ, Steingrímsdóttir ÓA, Stubhaug A, Nielsen CS. Sleep and pain sensitivity in adults. PAIN 2015;156:1433–9.
[78] Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 1998;17:87–97.
[79] Smallwood RF, Laird AR, Ramage AE, Parkinson AL, Lewis J, Clauw DJ, Williams DA, Schmidt-Wilcke T, Farrell MJ, Eickhoff SB, Robin DA. Structural brain anomalies and chronic pain: a quantitative meta-analysis of gray matter volume. J Pain 2013;14:663–75.
[80] Smith ML, Asada N, Malenka RC. Anterior cingulate inputs to nucleus accumbens control the social transfer of pain and analgesia. Science 2021;371:153–9.
[81] Spielberger CD. State-trait anxiety inventory for adults. APA SPsycTests, 1983.
[82] Spisak T. Statistical quantification of confounding bias in predictive modelling. arXiv 2021. preprint arXiv:211100814.
[83] Spisak T, Bingel U, Wager TD. Multivariate BWAS can be replicable with moderate sample sizes. Nature 2023;615:E4-E7.
[84] Spisak T, Kincses B, Schlitt F, Zunhammer M, Schmidt-Wilcke T, Kincses ZT, Bingel U. Pain-free resting-state functional brain connectivity predicts individual pain sensitivity. Nat Commun 2020;11:1–12.
[85] Spisák T, Pozsgay Z, Aranyi C, Dávid S, Kocsis P, Nyitrai G, Gajári D, Emri M, Czurkó A, Kincses ZT. Central sensitization-related changes of effective and functional connectivity in the rat inflammatory trigeminal pain model. Neuroscience 2017;344:133–47.
[86] Sullivan MJ, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychol Assess 1995;7:524–32.
[87] Talairach J, Tournoux P. Co-planar Stereotaxic Atlas of the Human Brain: 3-dimensional Proportional System: an Approach to Cerebral Imaging. Tieme, 1988.
[88] Talbot JD, Marrett S, Evans AC, Meyer E, Bushnell MC, Duncan GH. Multiple representations of pain in human cerebral cortex. Science 1991;251:1355–8.
[89] Tian Y, Zalesky A. Machine learning prediction of cognition from functional connectivity: are feature weights reliable? Neuroimage 2021;245:118648.
[90] Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Ser B (Methodological) 1996;58:267–88.
[91] Tracey I. Neuroimaging enters the pain biomarker arena. Sci Transl Med 2021;13:eabj7358.
[92] Tracey I, Woolf CJ, Andrews NA. Composite pain biomarker signatures for objective assessment and effective treatment. Neuron 2019;101:783–800.
[93] Tu Y, Zhang B, Cao J, Wilson G, Zhang Z, Kong J. Identifying inter-individual differences in pain threshold using brain connectome: a test-retest reproducible study. Neuroimage 2019;202:116049.
[94] Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TE, Bucholz R, Chang A, Chen L, Corbetta M, Curtiss SW, Della Penna S, Feinberg D, Glasser M, Harel N, Heath A, Larson-Prior L, Marcus D, Michalareas G, Moeller S, Oostenveld R, Petersen S, Prior F, Schlaggar B, Smith S,

Snyder A, Xu J, Yacoub E. The Human Connectome Project: a data acquisition perspective. Neuroimage 2012;62:2222–31.
[95] Varoquaux G. Cross-validation failure: small sample sizes lead to large error bars. Neuroimage 2018;180:68–77.
[96] Veldhuijzen DS, Nemenov MI, Keaser M, Zhuo J, Gullapalli RP, Greenspan JD. Differential brain activation associated with laser-evoked burning and pricking pain: an event-related fMRI study. PAIN 2009;141:104–13.
[97] Vidal-Pineiro D, Parker N, Shin J, French L, Grydeland H, Jackowski A, Mowinckel A, Patel Y, Pausova Z, Salum G, Sørensen Ø, Walhovd KB, Paus T, Fjell AM; Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing. Cellular correlates of cortical thinning throughout the lifespan. Scientific Rep 2020;10:21803.
[98] Voelkl B, Würbel H, Krzywinski M, Altman N. The standardization fallacy. Nat Methods 2021;18:5–7.
[99] Wager TD, Atlas LY, Lindquist MA, Roy M, Woo C-W, Kross E. An fMRI-based neurologic signature of physical pain. N Engl J Med 2013;368:1388–97.
[100] Wagstyl K, Ronan L, Goodyer IM, Fletcher PC. Cortical thickness gradients in structural hierarchies. Neuroimage 2015;111:241–50.
[101] Whelan CD, Altmann A, Botía JA, Jahanshad N, Hibar DP, Absil J, Alhusaini S, Alvim MK, Auvinen P, Bartolini E, Bergo FPG, Bernardes T, Blackmon K, Braga B, Caligiuri ME, Calvo A, Carr SJ, Chen J, Chen S, Cherubini A, David P, Domin M, Foley S, França W, Haaker G, Isaev D, Keller SS, Kotikalapudi R, Kowalczyk MA, Kuzniecky R, Langner S, Lenge M, Leyden KM, Liu M, Loi RQ, Martin P, Mascalchi M, Morita ME, Pariente JC, Rodríguez-Cruces R, Rummel C, Saavalainen T, Semmelroch MK, Severino M, Thomas RH, Tondelli M, Tortora D, Vaudano AE, Vivash L, von Podewils F, Wagner J, Weber B, Yao Y, Yasuda CL, Zhang G, Bargalló N, Bender B, Bernasconi N, Bernasconi A, Bernhardt BC, Blümcke I, Carlson C, Cavalleri GL, Cendes F, Concha L, Delanty N, Depondt C, Devinsky O, Doherty CP, Focke NK, Gambardella A, Guerrini R, Hamandi K, Jackson GD, Kälviäinen R, Kochunov P, Kwan P, Labate A, McDonald CR, Meletti S, O'Brien TJ, Ourselin S, Richardson MP, Striano P, Thesen T, Wiest R, Zhang J, Vezzani A, Ryten M, Thompson PM, Sisodiya SM. Structural brain abnormalities in the common epilepsies assessed in a worldwide ENIGMA study. Brain 2018;141:391–408.
[102] Woo C-W, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. Nat Neurosci 2017;20:365–77.
[103] Zhang X, Chen Q, Su Y, Meng J, Qiu J, Zheng W. Pain in the default mode network: a voxel-based morphometry study on thermal pain sensitivity. Neuroreport 2020;31:1030–5.
[104] Zou R, Li L, Zhang L, Huang G, Liang Z, Zhang Z. Predicting individual pain thresholds from morphological connectivity using structural MRI: a multivariate analysis study. Front Neurosci 2021;15:615944.
[105] Zunhammer M, Bingel U, Wager TD, Consortium PI. Placebo effects on the neurologic pain signature: a meta-analysis of individual participant functional magnetic resonance imaging data. JAMA Neurol 2018;75:1321–30.
[106] Zunhammer M, Schweizer LM, Witte V, Harris RE, Bingel U, Schmidt-Wilcke T. Combined glutamate and glutamine levels in pain-processing brain regions are associated with individual pain sensitivity. PAIN 2016;157:2248–56.