A replicable and generalizable neuroimaging-based indicator of pain sensitivity across individuals

Authors: Li-Bo Zhang[1,2,3], Xue-Jing Lu[1,2], Hui-Juan Zhang[1,2], Zhao-Xing Wei[2,4], Ya-Zhuo Kong[2,4], Yi-Heng Tu[1,2], Gian Domenico Iannetti[3,5], Li Hu[1,2,*]

Affiliations:

[1]CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences; 100101, Beijing, China;

[2]Department of Psychology, University of Chinese Academy of Sciences; 100049, Beijing, China;

[3]Neuroscience and Behaviour Laboratory, Italian Institute of Technology, 00161, Rome, Italy;

[4]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, 100101, Beijing, China;

[5]Department of Neuroscience, Physiology and Pharmacology, University College London, WC1E 6BT, London, UK.

*Corresponding author. Email: huli@psych.ac.cn

GDI and LH shared senior authorship

Abstract

Developing neural indicators of pain sensitivity is crucial for revealing the neural basis of individual differences in pain and advancing individualized treatment of pain. However, it still remains elusive whether pain-evoked neural responses can encode pain sensitivity. To address this issue, we analyzed five large functional magnetic resonance imaging (fMRI) datasets (total N = 1010), where healthy participants received painful and nonpainful tactile, auditory, and visual stimuli, and different pain treatments, including placebo and transcutaneous electrical neural stimulation. We systematically (1) investigated the correlation between pain-evoked fMRI responses and pain sensitivity, (2) evaluated the correlation's replicability in independent datasets and generalizability across different types of pain, (3) examined whether the correlation between fMRI responses and sensory sensitivity is unique to pain, and (4) how sample sizes affect the relationship between fMRI responses and pain sensitivity. We found that, with a sufficiently large sample size, there were replicable and generalizable correlations between pain-evoked fMRI responses and pain sensitivity across individuals for laser heat, contact heat, and mechanical pains. Despite lacking pain selectivity, fMRI signals exhibited larger correlations with pain sensitivity than with tactile, auditory, and visual sensitivity. Importantly, we developed a machine learning model that could accurately predict not only pain sensitivity to laser heat, contact heat, and mechanical stimuli, but also pain relief from pain treatments. Notably, our findings were influenced considerably by sample sizes, requiring >200 for univariate correlation analysis to reveal the relationship between pain sensitivity and fMRI responses, and >150 for multivariate analysis to decode pain sensitivity with fMRI responses. Altogether, given an enormous sample size, we convincingly showed the validity to decode pain sensitivity and predict analgesic effects using pain-evoked fMRI responses, which holds significant clinical promise in tailoring individualized pain treatments.

Keywords: pain sensitivity, individual differences, functional magnetic resonance imaging (fMRI), sample size, machine learning, analgesic effects

Introduction

Pain sensitivity varies widely across individuals (Mogil, 2021). The same nociceptive stimulus may be intolerably painful for someone, yet barely perceivable by another. What neural activity underlies such dramatic variability of pain sensitivity? This is a fundamental question in pain neuroscience. Understanding this is important not only for objectively measuring pain sensitivity, but also for improving the diagnosis and treatment of clinically relevant pain. High pain sensitivity is a risk factor for chronic pain (Campbell et al., 2013; Nielsen et al., 2009; Werner et al., 2010) – a condition that afflicts more than 20% of adults (Fayaz et al., 2016; Johannes et al., 2010). Revealing the neural underpinnings and identifying neural indicators of pain sensitivity could therefore aid in screening high-risk individuals for chronic pain and developing individualized pain treatment strategies (Davis et al., 2020; Tracey, 2021; Woolf & Max, 2001).

Numerous previous studies have attempted to find neural indicators of pain sensitivity with non-invasive neuroimaging techniques like functional magnetic resonance imaging (fMRI) (Coghill et al., 2003; Hoeppli et al., 2022; Spisak et al., 2020; Tu et al., 2019; L.-B. Zhang et al., 2022). However, three key issues remain incompletely addressed as of now. First and foremost, it is still subject to heated debate whether pain-evoked brain activations can serve as neural indicators of interindividual pain sensitivity. Painful stimuli consistently activate a series of brain areas such as the primary somatosensory cortex (S1), secondary somatosensory cortex (S2), anterior cingulate cortex (ACC), insula, thalamus, and prefrontal cortex (PFC) (Apkarian et al., 2005; Xu et al., 2020). An early seminal study linked larger pain-evoked blood oxygen level dependent (BOLD) responses in these areas with higher pain sensitivity (Coghill et al., 2003). Yet, a recent study from the same lab with a larger sample size (N = 101) demonstrated that no single brain area significantly correlated with interindividual pain sensitivity (Hoeppli et al., 2022), which then ignited a new round of debate (Gim et al., 2023; Goltermann & Büchel, 2023). These discrepant findings in fMRI studies may represent a fresh example of the replication problem and power failure in empirical sciences (Baker, 2016; Button et al., 2013; Open Science Collaboration, 2015; Poldrack et al., 2017). One critical cause of non-replicability is sample size (Bossier et al., 2020). In neuroimaging studies, a small sample size may be

insufficient to reliably detect the potential correlation between BOLD responses and behaviors at the interindividual level (Liu et al., 2023; Marek et al., 2022), especially when the true correlation is small (Gim et al., 2023; Schönbrodt & Perugini, 2013).

Another unresolved issue is whether pain-evoked brain responses selectively track pain sensitivity rather than modality-independent stimulus factors (Mouraux & Iannetti, 2018). Previous studies have shown that pain-evoked neural responses can be evoked by nonpainful but equally salient tactile, auditory, and visual stimuli (Hu & Iannetti, 2019; Mouraux et al., 2011; Ronga et al., 2012; Su et al., 2019). However, recent studies have found some small brain regions (Horing et al., 2019; X. Zhang et al., 2021) and brain activity patterns (Liang et al., 2019) that respond more to painful stimuli than to nonpainful stimuli, suggesting potentially preferential encoding of pain in the brain. Furthermore, rarely have these studies quantitatively examined whether certain brain region or activity pattern correlates with individual differences in pain and other modalities alike. As a result, previous studies have yet to provide a conclusive answer to the question whether there are selective neural indicators of interindividual pain sensitivity.

The third issue concerns the transferability of neural indicators of interindividual pain sensitivity. Such neural indicators have been primarily developed to assess pain sensitivity objectively (Mouraux & Iannetti, 2018; Su et al., 2020), as demonstrated by recent studies predicting pain sensitivity with MRI data (Kotikalapudi et al., 2022; Spisak et al., 2020; Tu et al., 2019). However, it has not been adequately evaluated whether neural indicators of pain sensitivity can be transferred to predict the outcomes of pain treatments, even though predicting medication responses is an oft-mentioned purpose of identifying neural indicators (Mouraux & Iannetti, 2018). Given the link between pain modulation system activity and pain sensitivity (Grouper et al., 2019; Skovbjerg et al., 2017), there is a plausible basis for neural indicators of pain sensitivity to be versatile enough to predict analgesic effects. It thus holds great potentials to investigate the transferability of neural indicators of pain sensitivity, which can foster individualized pain treatments.

To address these issues, we leveraged five large fMRI datasets (total N = 1010) where healthy participants received various types of pain stimuli: laser heat pain in Datasets 1 and 2 (L.-B. Zhang et al., 2022), mechanical pain in Dataset 3 (Botvinik-Nezer et al., 2023), contact heat pain in Datasets 4 (Gim et al., 2023) and 5 (Wei et al., 2024) (see Table S1 for a brief summary of each dataset). We aimed to answer four key questions (Figure 1A): (1) Do pain-evoked BOLD responses reflect pain sensitivity across individuals? (2) If so, is this correlation selective to the pain modality? (3) Can we develop a machine learning model that is versatile enough to accurately predict pain sensitivity and pain relief? (4) How many samples do we need to reliably decode pain sensitivity from BOLD responses? First, we correlated pain-evoked BOLD responses with pain sensitivity and tested the replicability and generalizability of the correlation with Datasets 1~3 (N = 794). Subsequently, we examined whether the correlation between BOLD responses and sensory sensitivity differ between pain and nonpain modalities with Datasets 1&2, where participants also received nonpainful tactile, auditory, and visual stimuli (Figure 1B). Next, we built a multivariate machine learning (ML) model, named the Neuroimaging-based Indicator of Pain Sensitivity (NIPS) model, to predict pain sensitivity using pain-evoked BOLD responses. We evaluated NIPS's performance with Datasets 1&2 and tested its generalizability across pain types and predictability of analgesic effects with three large external datasets, i.e., Datasets 3~5 (total N = 608). Finally, we systematically examined the influence of sample sizes on the ability to reliably detect the univariate correlation between BOLD responses and pain sensitivity and the performance of multivariate predictive models.

Results

BOLD responses reliably reflect pain sensitivity

We utilized Datasets 1~3 to examine whether pain-evoked BOLD activity reliably reflects inter-individual pain sensitivity. In Datasets 1&2, participants received and rated sensory stimuli of four modalities (painful, tactile, auditory, and visual) of two levels of physical intensity (low and high) (L.-B. Zhang et al., 2022); in Dataset 3, participants received mechanical pain stimuli of three levels of physical intensity (6kg/cm^2, 7kg/cm^2, and 8kg/cm^2) (Botvinik-Nezer et al., 2023). Dataset 1 delivered laser heat pain stimuli of 3.0J and 3.5J, and Dataset 2 delivered stimuli of 3.5J and 4.0J. To make full use of the large sample size, we pooled data from the 3.5J condition (N = 399) and tested the correlation between BOLD activity and pain sensitivity. We then assessed the replicability of the finding from this pooled dataset with data from the 3.0J condition in Dataset 1, the 4.0J condition in Dataset 2, and the mechanical pain data from Dataset 3. Finally, we explored the effect of sample sizes on our findings by generating bootstrapped samples from Datasets 1~3 (Wen et al., 2022). Note that Dataset 3 was originally collected to investigate placebo analgesia (Botvinik-Nezer et al., 2023). Here we utilized only the fMRI data in the control condition where no placebo cream was applied.

In Datasets 1&2, 10 laser heat pain stimuli per condition were delivered to participants. We quantified pain sensitivity as the mean pain rating for the 10 stimuli. As in previous studies (Gim et al., 2023; Hoeppli et al., 2022; Mogil, 2021), we observed considerable inter-individual variability of pain sensitivity even though the painful stimuli had identical physical intensity. Pain sensitivity ranged from 0.8 to 9.9 (mean ± standard deviation [SD]: 4.54 ± 1.72; Figure 2A and Table S4), covering almost the entire range of the 0 (‘‘no sensation’’) ~ 10 (‘‘the strongest sensation imaginable”) rating scale. Painful stimuli also activated a wide range of typically pain-related brain areas, including the S1, S2, ACC, insula, and thalamus (Figure 1C).

Importantly, we found significant Pearson’s correlations between pain-evoked BOLD signals and pain sensitivity across individuals (Figure 2A; p(FDR) < 0.05, the same hereafter). The majority of the correlations were positive, including the S1, S2, ACC, insula,

thalamus, precentral gyrus, supplementary motor area, precuneus, cerebellum, and the like. Negative correlations with pain sensitivity were only observed in a few areas, and the correlation coefficients were generally smaller than those of positive correlations were. To test the robustness of this finding, we also conducted Spearman's rank correlation analysis. A similar significant correlation pattern was found (Figure S1A; Pearson's r between the two unthresholded correlation maps = 0.958), demonstrating the robustness of our finding. In the analyses above, we accounted for head motion with the classical six motion parameters. To further control the potential influence of head motion, we regressed out from the BOLD signals the six motion parameters, their first derivatives, and squares of the motion parameters and first derivatives (Friston et al., 1996). Still, the significant correlation pattern resembled that of our original finding (Figure S1B; Pearson's r between the two unthresholded correlation maps = 0.936), suggesting a negligible influence of the head motion on our finding.

We then examined whether our finding could be replicated with other data. In both the 3.0J condition in Dataset 1 and the 4.0J condition in Dataset 2, significant correlations were present in a series of areas (Figure 2B). However, fewer areas showed significant correlations in these two conditions, which could be associated with their sample sizes being only half of the pooled 3.5J condition (~200 vs. ~400), thereby reducing the statistical power to detect true effects. In Dataset 3, we also found significant correlations between pain-evoked BOLD responses and pain sensitivity in all three physical intensity (i.e., 6kg/cm^2, 7kg/cm^2, and 8kg/cm^2) conditions (Figure 2C). Notably, fewer areas showed significant correlations as the physical intensity decreased. One plausible explanation could be that pain stimuli of low intensity, such as 6kg/cm^2, might not reliably induce pain sensations, potentially resulting in poor signal-to-noise ratio of the BOLD responses. Indeed, in the 6kg/cm^2 condition, 101 participants (27.2%) had a mean pain rating < 0.014, which was defined as "barely detectable pain". Consistent with laser heat pain in Datasets 1&2 and mechanical pain in Dataset 3, a previous study analyzing Dataset 4, where contact heat pain stimuli were delivered, also observed significant correlations between BOLD responses and pain sensitivity (Gim et al., 2023). Consistent findings in these datasets demonstrate not only the replicability of the correlation between BOLD responses and pain

sensitivity in different datasets, but also the generalizability of the correlation across different types of pain stimuli. To directly test whether classical pain-related areas encode pain sensitivity, we extracted mean signals from the S1, S2, ACC, insula, and thalamus in Datasets 1~3 using anatomically defined masks. Significant correlations were observed in most of these regions of interest (ROI) except when the physical intensity was low (i.e., 6kg/cm^2) in Dataset 3 (Table S5&S6), supporting that these pain-related areas indeed encode pain sensitivity.

Given the inconsistent findings in previous studies with smaller sample sizes (Coghill et al., 2003; Gim et al., 2023; Hoeppli et al., 2022), we systematically investigated the potential influence of sample size on the BOLD-pain sensitivity correlation in the 3.5J condition from Datasets 1&2. Briefly, we sampled with replacement 100~400 participants (in steps of 10) from Datasets 1&2, repeated the process 100 times, and estimated the sample size needed to detect a significant correlation in every voxel with a probability ≥ 0.8 (see Methods for details). The sample size needed for different voxels varied greatly across the brain (Figure 3A). Some areas only required a sample size of 100, while others needed more than 390. For voxels that showed significant correlations, the median sample size needed was 280, with an SD = 78 (Figure 3A). To specifically focus on the pain-related ROIs, we also estimated the probability of obtaining significant correlations for the S1, S2, ACC, insula, and thalamus (Figure 3B). With $\alpha = 0.01$, the S2 required the smallest sample size (≈ 150) to reach significance with a probability of 0.8; with $\alpha = 0.001$, the S2 required a sample size ≈ 220. Importantly, we also obtained similar findings in Dataset 3. In this independent dataset, a sample size of 270 was needed for half of the significant voxels to be detected with a probability ≥ 0.8 (Figure 3C). ROI analysis suggested that the insula required the smallest sample size of ≈ 180 and 240 to reach significance with a probability of 0.8, with $\alpha = 0.01$ and 0.001 respectively (Figure 3D). There were some differences in the exact sample size estimates and ROI analysis results between Datasets 1&2 and Dataset 3, possibly because different types of pain stimuli were employed. Nevertheless, these findings in independent datasets converge to reaffirm that the sample size had an enormous influence on the ability to detect significant correlations between BOLD signals and pain

sensitivity, and more than 200 participants would be need to reliably detect their correlations.

BOLD responses differentially reflect pain sensitivity and nonpain sensitivity

Having shown that BOLD responses reliably reflect pain sensitivity, we then examined in Datasets 1&2 whether the correlation is selective to the pain modality by correlating tactile, auditory, and visual stimuli-evoked BOLD responses with the corresponding sensory sensitivity. We also directly compared the correlation coefficients between modalities with Steiger's test (Steiger, 1980). Moreover, to rule out the potential confounding of rating differences between modalities, we manually selected participants who had no significant intensity rating differences between modalities (see Methods for details) and correlated BOLD signals with sensory sensitivity in these participants with matched pain and nonpain sensitivity.

In agreement with previous findings (Mouraux et al., 2011), brain regions activated by tactile, auditory, and visual stimuli were highly similar with those activated by pain stimuli (Figure 1C). However, Pearson's correlational results exhibited distinct patterns across modalities. BOLD responses evoked by high intensity tactile stimuli correlated significantly with tactile sensitivity in many brain areas (Figure 4A), while significant correlations only occurred in some small and scattered brain areas for high intensity auditory and visual stimuli (Figure 4B&C). These results remained largely the same if Spearman's rho was computed rather than Pearson's r (Figure S2). In the low intensity conditions, however, only a few brain areas showed significant positive correlations for all three nonpain modalities (Figure S3). As the stimulus intensity seemed to exert substantial influences on the results, we focused on the high intensity conditions hereafter. To avoid the imager's fallacy (Henson, 2005), we then directly compared correlation coefficients between pain and nonpain modalities. For pain and touch, significant differences in correlation coefficients were observed in some small areas, notably in the S1, precentral gyrus, and supplementary motor area (Figure 4D). For pain vs. audition and pain vs. vision comparisons, significant differences were found in much more areas, covering typical pain-related areas such as the S1, S2, ACC, insula, and thalamus (Figure 4D).

However, the foregoing findings might be confounded by the rating differences between modalities. Notably, there were significant differences in mean ratings among the four modalities ($F(3, 1194) = 117.295$, $p < 0.001$, partial $\eta^2 = 0.228$). Post-hoc tests revealed larger tactile ratings than pain and auditory ratings, and larger visual ratings than pain, tactile, and auditory ratings (Figure S4, Table S7). To equalize intensity ratings between modalities, we adopted a within-individual sensitivity matching approach, selecting participants whose 95% confidence interval of the mean difference between pain and nonpain ratings contained 0 (Tu et al., 2023). With this approach, we selected 192 participants for the pain-touch matching, 171 participants for the pain-audition matching, and 147 participants for the pain-vision matching. In the matched samples for pain-touch and pain-audition, we found that pain-evoked BOLD signals correlated with pain sensitivity in numerous areas (Figure 5A&B); in the matched samples for pain-vision, we only observed significant correlations in several small areas like the thalamus and supplementary motor area, possibly due to smaller sample sizes after pain-vision matching (Figure 5C). By contrast, in all three matched samples, BOLD responses in few, in any, areas showed significant correlations with tactile, auditory, and visual sensitivity (Figure 5A~C). We also directly compared correlation coefficients between modalities in these matched samples. A diverse array of regions had greater correlations with pain sensitivity than with nonpain sensitivity (Figure 5D).

Altogether, these findings demonstrate that there are quantitative differences in the correlation between BOLD responses and pain vs. nonpain sensitivity. Although BOLD responses do not selectively encode pain sensitivity, there is a potential for them to play a preferential role in this encoding.

NIPS accurately predicts pain sensitivity

We next developed a neuroimaging-based ML model (i.e., NIPS) to predict pain sensitivity across individuals. We first split randomly the pooled 3.5J data in Datasets 1&2 into a Discovery Set (N = 199) and a Holdout Test Set (N = 200). NIPS was developed with data in the Discovery Set and tested in the Holdout Test Set. To ensure NIPS's generalizability,

we assessed its predictive power in external Dataset 3 (N = 395) and Dataset 4 (N = 124). As in the univariate analyses, we also examined the effect of sample sizes on NIPS's performance.

To build the model, we used the least absolute shrinkage and selection operator-based principal component regression (LASSO-PCR; Figure 6A). In the Discovery Set, NIPS accurately predicted pain sensitivity with its 5-fold cross-validated performance as follows: $r = 0.531$, $p < 0.0001$, $R^2 = 0.242$ (Figure 6B). NIPS also performed well in the Holdout Test Set: $r = 0.496$, $p < 0.0001$, $R^2 = 0.237$ (Figure 6B). Even though NIPS was built based on the data from the 3.5J condition in the Discovery Set, it was also able to predict the mean pain ratings in the 3.0J condition (N = 101; $r = 0.404$, $p < 0.0001$, $R^2 = 0.146$) and the 4.0J condition (N = 99; $r = 0.489$, $p < 0.0001$, $R^2 = 0.229$) in the Holdout Test Set. To compare the performance of NIPS with other well-established pain prediction models, we applied the Neurological Pain Signature (NPS) (Wager et al., 2013) and Stimulus Intensity Independent Pain Signature-1 (SIIPS1) (Woo et al., 2017) to the Holdout Test Set. Although NPS failed to significantly predict pain sensitivity ($r = -0.004$, $p = 0.952$), SIIPS1 had significant predictive power ($r = 0.386$, $p < 0.0001$) (Figure 6C). A simple model based on the pain-related activation pattern derived from a term-based meta-analysis with neurosynth (Neurosynth "pain" model) also predicted pain sensitivity in the Holdout Test Set ($r = 0.224$, $p = 0.001$) (Figure 6C). Notably, NPS, SIIPS, and Neurosynth "pain" all had significantly smaller predictive power compared with NIPS: (1) NIPS vs. NPS: $t = 5.680$, $p < 0.0001$; (2) NIPS vs. SIIPS: $t = 2.275$, $p = 0.024$; (3) NIPS vs. Neurosynth "pain": $t = 4.027$, $p < 0.0001$. Altogether, these findings demonstrate that NIPS can accurately predict pain sensitivity, explaining more than 20% of the variance of pain sensitivity across individuals.

NIPS was developed using BOLD responses evoked by laser heat pain. To examine its generalizability, we then used it to predict mechanical pain sensitivity in Dataset 3 and contact heat pain sensitivity in Dataset 4. In Dataset 3, NIPS significantly predicted mechanical pain sensitivity in all three conditions: (1) the 8kg/cm^2 condition: $r = 0.274$, $p < 0.0001$; (2) the 7kg/cm^2 condition: $r = 0.217$, $p < 0.0001$; (3) the 6kg/cm^2 condition: r =

0.292, $p < 0.0001$ (Figure 6D). Likewise, significant correlations between predicted values and real values were found in Dataset 4: $r = 0.323$, $p = 0.0003$ (Figure 6D). Interestingly, this correlation was close to what the original authors of Dataset 4 observed ($r = 0.252$) in their own study, where a specific model was built to predict contact heat pain sensitivity (Gim et al., 2023). These results suggest that NIPS is generalizable to other pain types apart from laser heat pain.

To examine the effect of sample size on the model performance, we also resampled data (training sample size: 20 to 300, in steps of 5) in the pooled 3.5J condition in Datasets 1&2. Note that 5-fold cross-validation was adopted to assess the model performance. Therefore, for a sample of 100 participants, the training sample size (the abscissa in Figure 6E) is 80. When the training sample size was smaller than or equal to 70 (i.e., total sample size ≈ 88), NIPS's performance was very poor ($R^2 < 0$) and varied dramatically. When the training sample size was around 100 (i.e., 95, 100, or 105), the R^2 of NIPS was approximately 0.06. Interestingly, this estimate was very close to what Gim et al (2023) found with a total sample size of 124 (training sample size = 99; model performance: $r = 0.252$, $R^2 = 0.064$). Nevertheless, with such a sample size, NIPS's performance still varied substantially. The performance stabilized only after the training sample size reached 120 (i.e., total N = 150) and then R^2 kept increasing with the training sample size. These findings suggest that a large sample size (> 150) is needed to build a well-performing and robust pain prediction model in the multivariate analysis.

NIPS is a distributed pattern related to pain processing and modulation

Having testing NIPS's performance, we then attempted to understand what it represents. We first used virtual lesion to examine which brain areas contributed to NIPS's predictive power (Kohoutová et al., 2020). We then contextualized and interpreted the weight map of NIPS with neurosynth (Yarkoni et al., 2011) and neuromaps (Markello et al., 2022). We finally tested NIPS's transferability to predict the analgesic effects of placebo in Dataset 3 and neuromodulation in Dataset 5 (N = 92).

The weight map of NIPS exhibited widespread positive and negative weights across the entire brain (Figure 7A, p(FDR) < 0.05 for all voxels passing the strict 0.5/-0.5 threshold). Notably, pain-related areas typically had positive weights, such as the ACC, insula, thalamus, and cerebellum, while the PFC and occipital cortex showed negative weights. To explore which areas were more important, we parcellated the brain according to the Automatic Anatomical Labeling template (Tzourio-Mazoyer et al., 2002), removed (i.e., "virtually lesioned") each area, and refitted the predictive model without the corresponding area. We found that none of the areas was practically indispensable for NIPS's performance (Figure 7B; see supplemental data file for complete results), suggesting that information about pain sensitivity is highly distributed across the brain.

To facilitate the biological assessment of NIPS, we then conducted a term-based meta-analytical decoding of its weight map with neurosynth implemented in the brainstats toolbox (Larivière et al., 2023; Yarkoni et al., 2011). The weight map was most highly correlated with pain-related terms such as "noxious", "nociceptive", "pain", and "painful" (Figure S6A). In our own Datasets 1&2, the weight map of NIPS was also highly correlated with the pain activation pattern (r = 0.717; Figure 1C). Therefore, the weight map of NIPS largely covaried with the activation pattern of pain, suggesting that NIPS could also reflect pain processing intraindividually. If so, NIPS should not only predict pain sensitivity across individuals, but also single trial pain ratings within individuals. To test this hypothesis, we applied NIPS to single trial data in the Holdout Test Set. We observed significant correlations between predicted single trial ratings and real single trial ratings (mean r = 0.294, t(197) = 5.656, p < 0.0001, Figure S6B), supporting the notion that NIPS contained information about pain processing at the within-individual level.

To further understand what NIPS represents, we explored its neurotransmitter receptor profiles with neuromaps (Markello et al., 2022). We only observed a significant spatial correlation between the weight map of NIPS and the spatial distribution of the mu-opioid receptor (MOR) (r = 0.405, p_{spin}(FDR) = 0.034; Figure 7C), indicating that NIPS may be associated with pain modulation due to the key role of MOR in pain modulation system (Zubieta et al., 2001). To specifically test the hypothesis and evaluate NIPS's

transferability, we then applied NIPS to predict analgesic effects using Datasets 3&5. Dataset 3 was originally collected to investigate placebo analgesia (Botvinik-Nezer et al., 2023). We previously utilized the fMRI data in the control condition in Dataset 3 to test the generalizability of NIPS (Figure 6E). We now subtracted BOLD responses and mean pain ratings in the placebo condition from the control condition to quantify the placebo effect. NIPS significantly predicted the magnitude of placebo analgesia: (1) the 8kg/cm^2 condition: $r = 0.205$, $p < 0.0001$; (2) the 7kg/cm^2 condition: $r = 0.167$, $p = 0.002$; (3) the 6kg/cm^2 condition: $r = 0.204$, $p < 0.0001$ (Figure 7D). Dataset 5 was collected to study the analgesic effect of transcutaneous electrical nerve stimulation (TENS) with a pretest-treatment-posttest design (Wei et al., 2024). To quantify the analgesic effect, we subtracted BOLD responses and mean pain ratings in the posttest from those in the pretest. We also found a significant correlation between the model predictions and real pain reductions ($r = 0.254$, $p = 0.015$; Figure 7D). To account for the influence of baseline pain ratings (control condition in Dataset 3 and pretest condition in Dataset 5), we also conducted partial correlation analyses controlling baseline pain ratings. Still, predicted analgesic effects were significantly correlated with real pain reductions: (1) the 8kg/cm^2 condition in Dataset 3: partial $r = 0.181$, $p = 0.0005$; (2) the 7kg/cm^2 condition in Dataset 3: partial $r = 0.131$, $p = 0.015$; (3) the 6kg/cm^2 condition in Dataset 3: partial $r = 0.107$, $p = 0.041$; (3) Dataset 5: partial $r = 0.238$, $p = 0.023$. These findings converged to support the proposition that NIPS not only predict pain sensitivity, but also analgesic effects from pain treatments.

Discussion

With five large fMRI data (total N = 1010), we systematically examined the relationship between pain-evoked BOLD responses and pain sensitivity across individuals. We obtained four main findings. First, there were replicable and generalizable correlations between pain-evoked BOLD responses and inter-individual pain sensitivity in multiple large datasets. Second, the correlations between BOLD responses and sensory sensitivity were more evident in pain than in other modalities. Third, NIPS accurately predicted not only pain sensitivity across different pain types, but also pain relief from different pain treatments. Fourth, sample sizes had enormous impacts on not only the ability to reliably detect the univariate correlation between BOLD responses and pain sensitivity, but also the performance of the multivariate predictive model. Taken together, we compellingly demonstrate that given sufficiently large sample sizes, pain sensitivity can be decoded from pain-evoked fMRI responses, and a transferable neuroimaging-based ML model can be developed to predict pain sensitivity. These findings would have profound clinical significance to help foster individualized pain treatment and predict analgesic effects.

Due to its theoretical and clinical implications, understanding the neural basis of variability of pain sensitivity has been a major topic in pain neuroscience (Coghill et al., 2003; Hoeppli et al., 2022; Hu & Iannetti, 2019; Kotikalapudi et al., 2022; Melum et al., 2023; E. Schulz et al., 2012; Spisak et al., 2020; Tu et al., 2019; L.-B. Zhang et al., 2022). Theoretically, pain-evoked BOLD responses reflect real time neural processing of painful stimuli, thus potentially having a stronger correlation with pain sensitivity than other brain features, such as grey matter volume (Melum et al., 2023), cortical thickness (Kotikalapudi et al., 2022), and resting-state connectivity (Spisak et al., 2020; Tu et al., 2019). Indeed, task fMRI data do improve prediction of cognitive traits compared with resting-state fMRI data (Finn, 2021; Greene et al., 2018), and previous task fMRI studies suggested that pain-evoked BOLD responses could be considered as a promising neural correlate of pain sensitivity (Coghill et al., 2003; Gim et al., 2023). However, it is still highly controversial whether pain-evoked BOLD responses can reliably encode pain sensitivity, since apparently inconsistent findings were obtained in different groups even when the sample size (≈ 100) seemed large (Gim et al., 2023; Hoeppli et al., 2022). Leveraging the largest

datasets available to date, we found that pain-evoked BOLD responses did correlate with pain sensitivity. This correlation was robust regardless of whether parametric or nonparametric statistical methods were employed, and also replicable in different datasets with different stimulus intensities and generalizable across different pain types (i.e., laser heat pain, contact heat pain, and mechanical pain; Figure 2). In agreement with previous studies (Apkarian et al., 2005; Coghill et al., 2003; Gim et al., 2023), typical pain-related areas were also among the areas encoding pain sensitivity. We thus provided compelling evidence that, given sufficiently large sample sizes, pain sensitivity could be decoded from BOLD responses to pain, which is consistent in principle with previous studies showing the ability of pain-evoked electrophysiological responses to encode pain sensitivity (Hu & Iannetti, 2019; E. Schulz et al., 2012)

Having shown that BOLD responses reflect pain sensitivity, we adopted machine learning to develop NIPS, a model that predicts pain sensitivity from pain-evoked BOLD responses. With multiple independent datasets, we confirmed that NIPS has good predictive performance. Cross-validation and testing in the Holdout Test Set demonstrated that NIPS explained more than 20% of the variance of pain sensitivity (Figure 6). Directly pitting NIPS against other well-established models further showed NIPS seemed to have comparable or even higher predictive power (Wager et al., 2013; Woo et al., 2017). This observation may be explained by the fact that NIPS was specifically developed to prediction pain sensitivity, while others were not. Crucially, we tested NIPS's generalizability by applying it to mechanical pain data in Dataset 3 and contact heat pain data in Dataset 4. Although developed with laser heat pain data, NIPS exhibited significant predictive power in both datasets, demonstrating remarkable generalizability across different pain types. Note that contact heat pain and mechanical pain lasted for about 10 seconds in Datasets 3&4 and are believed to be more closely related to clinical pain compared with transient laser heat pain (J.-J. Lee et al., 2021). NIPS thus seemed to predict transient and sustained pain alike, further demonstrating its generalizability and potential applicability in clinical settings.

Given these robust and generalizable findings, why did some studies fail to find any correlation between BOLD responses and pain sensitivity? We examined two likely explanations. One is that the physical intensity of painful stimuli may be too small to evoke clear pain sensation (Goltermann & Büchel, 2023). The power to detect correlations will be dampened if no clear pain sensations are elicited or BOLD signals are just too noisy due to insufficient physical intensity. Our results from Dataset 3 supported this explanation: fewer areas showed significant correlations with pain sensitivity as the physical intensity decreased (Figure 2C). The other explanation is that the sample size may be simply not large enough. Traditionally, neuroimaging studies have limited sample sizes (Szucs & Ioannidis, 2020; Yeung, 2018), which is a crucial factor in the non-replicability of study findings (Bossier et al., 2020; Marek et al., 2022). Taking advantage of our large datasets, we systematically investigated the impact of sample size on the detectability of univariate correlations between pain-evoked BOLD signals and pain sensitivity with resampling analysis. We found that a sample size of around 200 could only detect less than 25% of the significant voxels (Figure 3). Similarly, sample sizes also had substantial influence on the predictive power of NIPS. Its predictive performance improved continuously as the sample size grew, suggesting that better models can be built with even more data (Cui & Gong, 2018; D. H. Lee et al., 2023; M.-A. Schulz et al., 2024). On the other hand, the performance of NIPS varied dramatically when the sample size was small. Only when the sample size reached 150 did we observed stable predictive performance. Therefore, a smaller sample size seems insufficient to reliably detect the correlations for both univariate and multivariate analyses. This conclusion resonates with recent studies showing that a massive sample size like several hundred or even thousand is required to detect replicable brain-behavior associations (Liu et al., 2023; Marek et al., 2022; Spisak et al., 2023; Turner et al., 2018).

Although pain-evoked BOLD responses reliably correlated with pain sensitivity, this correlation seemed not very pain-selective. As in previous studies (Mouraux et al., 2011; Su et al., 2019; Tu et al., 2023), pain and nonpain stimuli activated more or less the same brain regions in our study (Figure 1C). Likewise, we also found significant correlations between nonpain stimuli-evoked BOLD responses and corresponding sensory sensitivity

(Figure 4A~C), which is consistent with a recent study showing correlations between BOLD signals and auditory sensitivity (Hoeppli et al., 2022). However, we observed larger correlations between BOLD signals and pain sensitivity than between BOLD signals and tactile/auditory/visual sensitivity (Figure 4D). Importantly, similar findings were obtained when sensitivity differences between modalities were controlled by a matching procedure (Figure 5). Therefore, although the correlation between BOLD signals and sensory sensitivity is not specific to pain, there are quantitatively larger correlations in pain than in other sensory modalities. In other words, the brain is more capable of representing pain sensitivity, showing a potentially preferential role in encoding pain sensitivity. Future research is in need to reveal the underlying mechanisms of differential encoding of pain sensitivity and nonpain sensitivity.

Apart from establishing the correlation between BOLD responses and pain sensitivity, the present study also deepened our understanding of neural representations of pain sensitivity and suggested a promising clinical application of NIPS. When removing brain areas one by one, we observed few to no reductions in NIPS's predictive power. This finding is in agreement with real brain lesion studies showing that lesions in key pain-related areas do not disrupt pain intensity perception (Knecht et al., 1996; Starr et al., 2009). Recent studies also demonstrated that pain can be decoded from nearly all brain areas (Vigotsky et al., 2022). Altogether, our observation supported the notion that pain is represented in a distributed neural system, rather than local and isolated regions (Coghill, 2020).

Crucially, the weight map of NIPS significantly correlated with the spatial distribution of MOR. Since MOR is heavily involved in pain modulation (Zubieta et al., 2001), we hypothesized that NIPS could be related to pain modulation as well. Directly applying NIPS to Datasets 3&5 confirmed this hypothesis: NIPS could be transferred to significantly predict pain reduction in both datasets with different pain treatments. These findings bridge two types of individual differences, namely, pain sensitivity and pain relief, suggesting that pain sensitivity may be co-determined by individuals' ability to both process and modulate nociceptive information. This statement was supported by the difference in the power in predicting pain sensitivity and analgesic effects: NIPS explained more than 20% of the

variance of pain sensitivity but less than 10% of the variance of analgesic effects. Notably, consistent with this notion, some studies discovered significant associations between pain sensitivity and conditioned pain modulation (Grouper et al., 2019; Skovbjerg et al., 2017). In short, our findings imply that NIPS can be applied to predict analgesic effects of distinct pain treatments, thus suggesting a direct clinical implication of NIPS.

Despite the theoretical and clinical significance of our findings, it is important to acknowledge some limitations. First, we only investigated the influence of stimulus intensity and sample size on the detectability of the correlation between BOLD responses and pain sensitivity due to the nature of datasets available. Other factors like sample personal characteristics and study design may also contribute to explaining the variability of pain sensitivity, which should be tested in future studies. Second, while NIPS has reasonably good performance in predicting pain sensitivity, its predictive power can be further improved. As suggested by our resampling analysis, collecting even more data has the potential to boost NIPS's predictive performance (M.-A. Schulz et al., 2020). Incorporating multimodal data, such as resting-state fMRI and structural MRI, in NIPS may be helpful to capture more variability of pain sensitivity, thus further improving the predictive performance (M.-A. Schulz et al., 2024). Finally, participants in all datasets were healthy individuals. It thus remains to be tested whether our findings can be generalized to chronic pain patients, and it is of particular interest to test whether NIPS can predict pain relief in patient groups.

Methods

Datasets overview

We unitized five large fMRI datasets that had been published previously to study other research questions (Figure 1A). We collected Datasets 1, 2, and 5 on our own. Datasets 3 and 4 were kindly shared by the original authors (Botvinik-Nezer et al., 2023; Gim et al., 2023). In Datasets 1 and 2, participants received laser pain, electro-tactile, auditory, and visual stimuli (L.-B. Zhang et al., 2022). Dataset 3 were originally collected to investigate the neural underpinnings of placebo analgesia (Botvinik-Nezer et al., 2023). The researchers delivered both contact heat pain and mechanical pain stimuli to participants, but used contact heat pain stimuli to induce placebo analgesia by conditioning. To avoid potential influence of the placebo induction on the neural processing of contact heat pain, we analyzed data only for the mechanical pain stimuli in the placebo and control conditions, and discarded data with contact heat pain stimuli. Instead, we used data with contact heat pain stimuli from Dataset 4 (Gim et al., 2023), where no additional interventions like placebo were introduced. Dataset 5 also used contact heat pain stimuli, and was from a study aiming to reveal the spino-cortical mechanisms of TENS-induced analgesia (Wei et al., 2024). In Datasets 1~4, the physical intensity of pain stimuli was fixed. In Dataset 5, however, the physical intensity was adjusted individually for each participant. Therefore, we did not include this dataset in most analyses, but only applied NIPS to it to test whether NIPS could predict pain relief. We summarized basic information for these five datasets in Table S1.

Participants

Dataset 1 included 212 participants. One participant did not provide necessary demographic information. For the rest 211 participants, there were 135 females, who aged 21.5 ± 4.2 years (Mean ± SD). Dataset 2 had 187 participants. There was also one participant who did not provide the demographic information. For the rest 186 participants, there were 103 females, aging 21.0 ± 3.3 years. Dataset 3 included 395 participants (231 females, age: 35.4 ± 2.6). Note that some participants in this dataset had missing behavioral ratings in some conditions. Considering the lack of sufficient trials per condition, we discarded data from these participants with missing values, which was the reason why the

sample size in the main figures was not 395. Dataset 4 had 124 participants (61 females, age: 22.2 ± 2.7). Dataset 5 included 92 participants (50 females, age: 21.9 ± 3.2). All participants were healthy and free of pain, neurological, or psychiatric disorders. They all gave written informed consent prior to the experiments. Local ethics committees approved the experimental procedures for the original studies (Datasets 1, 2, and 5: Ethics Committee of the Institute of Psychology at the Chinese Academy of Science; Dataset 3: Institutional Review Board of the University of Colorado Boulder; Dataset 4: Sungkyunkwan University Institutional Review Board).

Sensory stimulation

Three different types of pain stimuli were used in these datasets: (1) laser heat pain: Datasets 1 and 2; (2) mechanical pain: Dataset 3; (3) contact heat pain: Datasets 4 and 5. Datasets 1 and 2 also delivered three kinds of nonpain stimuli: (1) electro-tactile, (2) auditory, and (3) visual stimuli.

In Datasets 1 and 2, laser heat pain stimuli were transient radiant heat pulses (wavelength: 1.34μm; pulse duration: 4ms) generated by an infrared neodymium yttrium aluminum perovskite (Nd: YAP) laser (Electronical Engineering, Italy). The laser beam was transmitted by an optic fiber, and its diameter was set at approximately 7mm. Laser pulses were delivered to a pre-defined square ($5\times5cm^2$) on the left hand dorsum. After each stimulus, the laser beam was displaced by approximately 1cm in a random direction to avoid nociceptor fatigue or sensitization. Two stimulus energies (3.0J and 3.5J) were used in Dataset 1, and two other stimulus energies (3.5J and 4.0J) were used in Dataset 2. In Dataset 3, mechanical pain stimuli were delivered to the fingers of the left hand for 10s with an MRI-compatible pressure pain device. There were three levels of physical intensity, 6 kg/cm^2, 7 kg/cm^2, and 8kg/cm^2. In Datasets 4, contact heat pain stimuli were administered to the volar surface of the left forearm with a $16\times16mm^2$ ATS thermode of a Pathway system (Medoc Ltd, Ramat Yishai, Israel). The target temperature of the stimuli was 47.5°C with the baseline temperature being 32°C. The stimulation lasted for 12s (ramp-up: 2.5s; plateau: 7s; ramp-down: 2.5s). In Dataset 5, contact heat pain stimuli were delivered to the C5~C6 dermatomes on both forearms with a $573mm^2$ (diameter of 27mm) CHEPS

thermode using a Pathway system (Medoc Ltd, Ramat Yishai, Israel). The heat stimulation ramped up from a baseline of 32°C to a temperature for 3s that led to a perceived intensity rating of 7 on a 0~10 rating scale. In other words, the target temperature was tailored individually for each participant to ensure a constant perceived intensity.

In Datasets 1 and 2, nonpain stimuli were also delivered. Non-nociceptive tactile stimuli were constant current square-wave electrical pulses (duration: 1ms; model DS7A, Digitimer, UK) delivered through a pair of skin electrodes (1cm interelectrode distance) placed on the left wrist, over the superficial radial nerve. The same two stimulus intensities (2.0mA and 4.0mA) were used both datasets. Auditory stimuli were brief pure tones (frequency: 800Hz; duration: 50ms; 5ms rise and fall time) delivered through a headphone. The same two stimulus intensities (76dB SPL and 88dB SPL) were used for all participants in these datasets. Visual stimuli were brief flashes of a grey round disk in a black background (duration: 100ms) on a computer screen. The stimulus intensities were adjusted using the greyscale of the round disk, corresponding to RGB values of (100, 100, 100) and (200, 200, 200), respectively. Stimulus intensities of tactile, auditory, and visual stimuli were determined based on a pilot behavioral experiment to ensure that the perceived ratings of low and high intensity stimuli were approximately 4 and 7 out of 10, respectively.

Study design

The original authors reported their study designs in great detail (Botvinik-Nezer et al., 2023; Gim et al., 2024; Wei et al., 2024; L.-B. Zhang et al., 2022). Since these papers were all openly accessible, here we only briefly described the study design of each dataset (Figure 1B).

In Datasets 1 and 2, participants received 80 transient stimuli of four different sensory modalities (nociceptive laser, non-nociceptive tactile, auditory, and visual) divided into two runs, and then rated their perceived intensity with a rating scale ranging from ranging from 0 (‘‘no sensation’’) to 10 (‘‘the strongest sensation imaginable [in each stimulus modality]’’). For each sensory modality, two stimulus intensities (i.e., high and low) were delivered. In other words, each participant underwent eight experimental conditions (4

modalities×2 intensities), each condition with 10 trials. The only difference in the study design of Datasets 1 and 2 was the physical intensity of laser heat pain stimuli. In Dataset 1, stimuli of 3.0J and 3.5J were delivered, while in Dataset 2 stimuli of 3.5J and 4.0J were administered. A stimulus intensity of 3.0J was adopted in Dataset 1 because nociceptive laser stimuli of 4.0J were unbearable for these participants

Dataset 3 adopted a within-subjects design to examine the neural mechanisms of placebo analgesia. The study consisted of a verbal suggestion phase, a conditioning phase, and a test phase. Verbal suggestions and conditioning were used to induce placebo effects. In the verbal suggestion phase, two identical creams were introduced to participants, one being a placebo cream that was claimed to have pain-relieving effect, the other being a control cream claimed to have no analgesic effects. These two creams were applied to two different fingers of the left hand. Then participants went through the conditioning phase, where they experienced the analgesic effect of the placebo cream through social observational learning and classical conditioning. Importantly, in this conditioning phase, only contact heat pain stimuli were delivered. Finally, in the test phase, participants received both contact heat pain and mechanical pain stimuli in the fingers with the placebo and control creams. The test phase delivered a total of 32 stimuli in four runs, with four thermal and four mechanical stimuli in a random order per run. After each stimulus, participants rated the intensity and unpleasantness of the stimulus with a 0~1 rating scale, where 0 means “no pain”, 0.014 “barely detectable”, 0.061 “weak pain”, 0.172 “moderate pain”, 0.354 “strong pain”, and 1 “most intense sensation imaginable”.

In Dataset 4, participants received contact heat pain stimuli of six intensity levels in their left forearm. The entire study consisted of eight pain task runs with 12 trials each run and 16 trials per temperature. After each stimulus, participants rated the perceived intensity of the stimuli with a 0~1 rating scale (0, “no sensation”; 0.061, “weak”; 0.172, “moderate”; 0.3454, “strong”; 0.533, “very strong”; 1, “strongest imaginable”). Note that the original authors only shared data for the highest temperature condition (i.e., 47.5°C). We thus could only analyze data for this condition.

Different from Datasets 1~4, Dataset 5 adopted a between-subjects design to investigate the cortico-spinal mechanisms of TENS-induced analgesia. Participants were randomly assigned to three groups: conventional TENS (c-TENS), acupuncture-like TENS (a-TENS), and sham TENS. Each participant experienced 60 contact heat pain stimuli (pre-treatment session: 15 stimuli in the left arm and 15 stimuli in the right arm; post-treatment session: 15 stimuli in the left arm and 15 stimuli in the right arm) in two runs and then reported their perceived intensity and unpleasantness in a 0~10 ratings scale (0: "no feeling/unpleasantness"; 10: "unbearable pain/extreme unpleasantness"). Note that we only used data from the 15 stimuli delivered to the left arm, since pain stimuli were only delivered to the left hand in other datasets. We pooled data from all three groups. The purpose was twofold. First, we could have a large sample size and thus a larger statistical power. Second, our goal was to test whether NIPS could predict the amount of pain reduction after the treatment compared with the pre-treatment baseline, but not the analgesic effect of verum TENS compared with sham TENS. Indeed, pain sensitivity could also change after the sham TENS treatment due to factors like natural history, regression to the mean, and so on. A model that predicts pain reduction should thus also predict sensitivity change in the sham TENS condition.

MRI acquisition

Details of the MRI acquisition parameters can be found in Table S2&S3.

Image preprocessing

We analyzed preprocessed data from the original studies (Botvinik-Nezer et al., 2023; Gim et al., 2023; Wei et al., 2024; L.-B. Zhang et al., 2022). Detailed preprocessing descriptions can be found in these papers. For Datasets 1&2, fMRI data were preprocessed using Statistical Parametric Mapping 12 (SPM12) (Wellcome Trust Center for Neuroimaging, London). The preprocessing included: (1) removing the first three volumes in each run; (2) slice-time correction using the second slice and realignment to the mean slice; (3) co-registration; (4) normalizing to the Montreal Neurological Institute (MNI) space (resampling voxel size = $3\times3\times3mm^3$); (5) regressing out five principal components of the

white matter (WM) and cerebrospinal fluid (CSF) signals, and the six motion parameters; (6) smoothing with a 6mm full-width at half maximum (FWHM) Gaussian kernel.

For Dataset 3, images were preprocessed with fMRIprep ver. 20.2.3 (Esteban et al., 2019). Briefly, the preprocessing included: (1) motion-correction with MCFLIRT (FSL 5.0.9); (2) co-registering to the T1 reference using the boundary-based registration method (BBR) (FreeSurfer); (3) normalizing to the standard space; (4) smoothing with a 6mm FWHM Gaussian kernel.

For Dataset 4, preprocessing was performed with SPM12 and FSL, including: (1) removing the first 18 volumes for each run; (2) realignment; (3) co-registration; (4) normalizing to the MNI space (2×2×2 mm^3); (5) smoothing with a 5mm FWHM Gaussian kernel; (6) reducing motion-related artefacts with the Independent Component Analysis-based strategy for Automatic Removal Of Motion Artifacts; (7) removing data where the mean frame displacement (FD) of a run > 0.2mm, or the FD of any volume > 5mm.

For Dataset 5, preprocessing was also conducted using FSL, with the following steps: (1) physiological noise (i.e., respiratory and cardiac noise) regression with the physiological noise model toolbox; (2) distortion correction with the TOPUP tool; (3) slice timing correction with the *slicetimer* function; (4) motion-correction with MCFLIRT; (5) brain extraction with the BET; (6) smoothing with a 5mm FWHM Gaussian kernel; (7) high-pass filtering (cutoff: 100s); (8) co-registration with the BBR; (9) normalizing to the MNI space (2×2×2 mm^3) with FNIRT.

fMRI data analysis

First level analysis

We mainly reanalyzed 1st level contrast images from the original studies (Botvinik-Nezer et al., 2023; Gim et al., 2023; Wei et al., 2024; L.-B. Zhang et al., 2022). In Datasets 1&2, regressors in the 1st level analysis included eight conditions (4 modalities×2 intensities) convolved with the canonical hemodynamic response function, its temporal derivatives, and six head motion estimates. Moreover, we high-pass filtered images with a cutoff period

of 128s and accounted for temporal autocorrelations using the first-order autoregressive model (AR(1)). In Dataset 3, regressors in the 1st level analysis included all experimental conditions convolved with the canonical hemodynamic response function, 24 motion parameters, and mean CSF signal. A high-pass filter of 180s was also applied. In Dataset 4, a single trial approach was utilized. Regressors included every single pain trial convolved with the canonical hemodynamic response function, five principal components of WM and CSF signal, and a linear trend. Signals were also filtered with a 180s high-pass filter. Trials with a variance inflation factor > 3 were subsequently removed. Single trial beta maps were finally averaged to derive the beta map for pain stimulation. In Dataset 5, regressors included pain stimulation and rating period convolved with the canonical hemodynamic response function and its time derivatives.

Second level correlational analysis

To examine the relationship between pain-evoked BOLD responses and pain sensitivity, we correlated 1st level beta maps in each pain condition with average pain ratings. The significance threshold was set at p(FDR) = 0.05 (two-tailed). To test the robustness of results, we computed both Pearson's r and Spearman's rho in the correlational analyses. In addition to whole brain analysis, we also conducted ROI analysis. Specifically, we extracted mean signals from classical pain-related areas (Apkarian et al., 2005; Xu et al., 2020), namely, the S1, S2, ACC, insula, and thalamus using anatomically defined masks defined with the Harvard-Oxford Atlas distributed with FSL (Desikan et al., 2006). Note that the Harvard-Oxford Atlas is a probabilistic atlas and we thresholded probability maps for each region at 50% (Geuter et al., 2017).

To test the pain selectivity of the correlation between BOLD responses and sensory sensitivity, we also performed the same correlational analysis for touch, audition, and vision in Datasets 1&2. To directly test the significance of correlation differences, we compared the correlation coefficients between modalities with Steiger's test (Steiger, 1980).

Sensitivity matching procedure

Although physical intensity of sensory stimuli was fixed in Datasets 1&2, perceived intensity ratings (i.e., sensory sensitivity) were significantly different among different modalities. Distinct correlational results in pain and nonpain modalities could thus be potentially attributed to intensity rating differences. Furthermore, perceived intensity ratings are highly correlated with stimulus salience (Liang et al., 2019). As a result, differences in perceptual ratings may result in incompatible salience levels between modalities. To control for this possible confounding factor of rating differences, we adopted a rating matching procedure to equalize intensity ratings between pain and nonpain modalities. To rule out the influence of individual differences, we matched pain intensity ratings and tactile/auditory/visual intensity ratings in a within-individual manner. Specifically, we manually screened participants whose pain intensity ratings in the 3.5J condition and tactile/auditory/visual intensity ratings in the high intensity condition were not significantly different, namely the 95% confidence interval of the rating differences contained 0. With this matching criterion, 192 participants were selected for the pain-touch comparison, 171 for the pain-audition comparison, and 147 for the pain-vision comparison (Figure 5A~C).

Resampling procedure

We used a resampling method to examine the influence of sample sizes on the probability of detecting significant correlations between BOLD responses and pain sensitivity. In the pooled 3.5J data from Datasets 1&2 (N = 399), we generated bootstrapped subsamples from the whole dataset, that is, randomly selected data samples with replacement. The sample size of these subsamples ranged from 100 to 400 (in steps of 10), and, for each sample size, the resampling procedure was repeated for 100 times. In each subsample, we correlated BOLD responses with pain sensitivity and thresholded the correlation maps with p(FDR) = 0.05. Since we generated 100 subsamples for each sample size, the probability of a voxel being significant under a certain sample size can be approximated as the number of times that voxel reached significance. We then calculated the sample size needed for each voxel to pass the p(FDR) = 0.05 threshold with a probability of 0.8.

To supplement this whole brain analysis, we also conducted the resampling in a ROI analysis manner. Specifically, we defined anatomical masks for the S1, S2, ACC, insula, and thalamus using the Harvard-Oxford Atlas (Desikan et al., 2006) distributed with FSL. Note that the probability maps for each region were thresholded at 50% (Geuter et al., 2017). Subsequently, we generated 1000 bootstrapped subsamples with sample sizes from 10 to 400 (in steps of 10). In this ROI analysis, we set the significance to $\alpha = 0.01$ or $\alpha = 0.001$, and then made line charts to fully show the relationship between sample sizes and the probability of significance for each ROI.

To test the robustness of the findings from the foregoing resampling method, we also conducted the whole brain and ROI analyses with data for 8kg/cm^2 condition in the Dataset 3 (N = 369). Note that the upper bound of the exampled sample size was set to 370 due to a slightly smaller sample size for Dataset 3.

Machine learning

Model development

We split randomly the pooled 3.5J data in Datasets 1&2 into a Discovery Set (N = 199) and a Holdout Test Set (N = 200). A LASSO-PCR model was trained to predict pain sensitivity using 1st level t maps in the Discovery Set with Scikit-learn (https://scikit-learn.org/stable/index.html). The LASSO is a regularization method that shrinks the linear regression coefficient estimates toward zero (Tibshirani, 1996). Mathematically, LASSO regression minimizes the following loss function:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p}|\beta_j|$$

where y_i is the predicted variable of participant i, β_0 and β_j are the linear regression coefficient estimates, x_{ij} is the j-th predictor for participant i, λ is the tuning parameter, n is the number of participants, and p is the number of predictors.

T maps were first standardized across participants and then submitted to principal component analysis to reduce the feature dimensions. Principal components were used as features in the LASSO regression model. Five-fold cross-validation was used to evaluate

model performance. The predictive performance of the final model was assessed using two metrics: (1) Pearson's correlation coefficient between the predicted and real pain sensitivity, and (2) R^2, which is close to the square of Pearson's correlation if the mean predictive error is 0 but can be negative if the mean predictive error is large. To determine the optimal tuning parameter λ, we selected 100 λ values log-uniformly distributed in the $[10^{-2}, 10^{2}]$ interval and tuned λ according to the cross-validated mean squared errors. The optimal λ was 0.236.

Model testing

After our model (i.e., NIPS) was developed, we tested its performance in the Holdout Test Set. To compare NIPS and other well-recognized models, we applied the Neurological Pain Signature (NPS) (Wager et al., 2013) and Stimulus Intensity Independent Pain Signature-1 (SIIPS1) (Woo et al., 2017) to the Holdout Test Set. In addition, we also built a simple model based on the pain-related activation pattern derived from a term-based meta-analysis with neurosynth (Neurosynth "pain" model) (Yarkoni et al., 2011). The predictive values (model expression) of these three models were the dot product of the weight maps of the models and 1st level t map. Pearson's correlation coefficients between predicted and real values were also directly compared to test if NIPS performed better than these three models. NIPS was developed with laser heat pain data in Datasets 1&2. To test its generalizability, we then applied the model to mechanical pain data in Dataset 3 and contact heat pain data in Dataset 4. Note that the original authors of Datasets 3&4 shared beta maps for pain, not t maps. NIPS's performance in these two datasets may thus be slightly affected due to scaling issues in beta maps.

To examine the influence of sample size on NIPS's performance, we refitted the model with different training sample size with the "learning_curve" function in Scikit-learn. The learning_curve function generates a curve showing how the performance of the model changes as the training sample size grows. We first randomly shuffled the pooled 3.5J data in Datasets 1&2, and then inputted the model and the shuffled dataset to the learning_curve function. We considered training sample sizes from 20 to 300 (in steps of 5). This whole

process was repeated for 100 times to also estimate the variability of the relationship between training sample size and model performance.

Model interpretation

We employed three approaches to interpret NIPS's weight maps: virtual lesion (Kohoutová et al., 2020), meta-analytical decoding based on neurosynth (Yarkoni et al., 2011), and neurotransmitter receptor profiling with neuromaps (Markello et al., 2022). For virtual lesion, we parcellated NIPS's weight map according to the Automatic Anatomical Labeling template (Tzourio-Mazoyer et al., 2002) and removed areas one by one to examine which area's removal had substantial impact on the model performance.

Moreover, we conducted a term-based meta-analytic decoding analysis with neurosynth implemented in the brainstats toolbox (Larivière et al., 2023; Yarkoni et al., 2011). Briefly, this analysis correlated the weight map of NIPS with activation maps derived from large-scale term-based meta-analysis. Since the weight map was mainly associated with pain-related terms, we hypothesized that NIPS could reflect pain processing. To empirically test this hypothesis, we also applied NIPS to predict single-trial pain ratings in the Holdout Test Set.

To reveal the neurotransmitter receptor profile of NIPS, we then correlated its weight map with the spatial distribution maps of major neurotransmitter receptors with the neuromaps toolbox (Markello et al., 2022). The neurotransmitter receptors we examined included: (1) alpha-4 beta-2 nicotinic receptor (a4b2) (Hillmer et al., 2016); (2) muscarinic acetylcholine receptor subtype M1 (Naganawa et al., 2021); (3) vesicular acetylcholine transporter (VAChT) (Aghourian et al., 2017); (4) cannabinoid type-1 receptor (CB1) (Normandin et al., 2015); (5) dopamine D1-type receptor (Kaller et al., 2017); (6) dopamine D2-type receptor (Sandiego et al., 2015); (7) respective dopamine transporter (DAT) (Dukart et al., 2018); (8) GABAa receptor (Nørgaard et al., 2021); (9) metabotropic glutamate type 5 receptor (mGluR5) (Smart et al., 2019); (10) histamine H3 receptors (Gallezot et al., 2017); (11) μ-opioid receptor (MOR) (Kantonen et al., 2020); (12) norepinephrine transporter (NET) (Ding et al., 2010); (13) serotonin 1A receptor (5-HT1a) (Savli et al., 2012); (14) 5-

HT1b (Beliveau et al., 2017); (15) 5-HT2a (Beliveau et al., 2017); (16) 5-HT4 (Beliveau et al., 2017); (17) 5-HT6 (Radhakrishnan et al., 2018); (18) serotonin transporter (5-HTT) (Beliveau et al., 2017). To determine the statistical significance, we used the spin test with 10000 permutations (Alexander-Bloch et al., 2018). Since the weight map of NIPS was correlated with MOR, NIPS could be associated with pain modulation. To empirically test this hypothesis and assess NIPS's transferability, we applied NIPS to Datasets 3&5 to predict their analgesic effects from different pain treatments.

Data and code availability

The data and code for all results will be made available on the Open Science Framework (OSF) upon publication.

Acknowledgement

We thank colleagues collecting Datasets 3&4 for kindly sharing their data. This work was supported by the National Key Research and Development Program of China (2023YFC2508702), National Natural Science Foundation of China (32071061), and Beijing Natural Science Foundation (JQ22018).

Author Contributions

Conceptualization: LBZ, LH; Methodology: LBZ, LH; Investigation: HJZ, LH, ZXW, LBZ; Formal analysis: LBZ, LH; Visualization: LBZ, LH; Writing – original draft: LBZ; Writing – review & editing: LBZ, XJL, HJZ, ZXW, YZK, YHT, GDI, LH; Funding acquisition: LH; Project administration: LH; Supervision: LH, GDI.

References

Aghourian, M., Legault-Denis, C., Soucy, J.-P., Rosa-Neto, P., Gauthier, S., Kostikov, A., Gravel, P., & Bédard, M.-A. (2017). Quantification of brain cholinergic denervation in Alzheimer's disease using PET imaging with [18F]-FEOBV. *Molecular Psychiatry*, *22*(11), 1531–1538. https://doi.org/10.1038/mp.2017.183

Alexander-Bloch, A. F., Shou, H., Liu, S., Satterthwaite, T. D., Glahn, D. C., Shinohara, R. T., Vandekar, S. N., & Raznahan, A. (2018). On testing for spatial correspondence between maps of human brain structure and function. *NeuroImage*, *178*, 540–551. https://doi.org/10.1016/j.neuroimage.2018.05.070

Apkarian, A. V., Bushnell, M. C., Treede, R.-D., & Zubieta, J.-K. (2005). Human brain mechanisms of pain perception and regulation in health and disease. *European Journal of Pain*, *9*(4), 463–463. https://doi.org/10.1016/j.ejpain.2004.11.001

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), Article 7604. https://doi.org/10.1038/533452a

Beliveau, V., Ganz, M., Feng, L., Ozenne, B., Højgaard, L., Fisher, P. M., Svarer, C., Greve, D. N., & Knudsen, G. M. (2017). A High-Resolution In Vivo Atlas of the Human Brain's Serotonin System. *Journal of Neuroscience*, *37*(1), 120–128. https://doi.org/10.1523/JNEUROSCI.2830-16.2016

Bossier, H., Roels, S. P., Seurinck, R., Banaschewski, T., Barker, G. J., Bokde, A. L. W., Quinlan, E. B., Desrivières, S., Flor, H., Grigis, A., Garavan, H., Gowland, P., Heinz, A., Ittermann, B., Martinot, J.-L., Artiges, E., Nees, F., Orfanos, D. P., Poustka, L., … Moerkerke, B. (2020). The empirical replicability of task-based fMRI as a function of sample size. *NeuroImage*, *212*, 116601. https://doi.org/10.1016/j.neuroimage.2020.116601

Botvinik-Nezer, R., Petre, B., Ceko, M., Lindquist, M. A., Friedman, N. P., & Wager, T. D. (2023). *Placebo treatment affects brain systems related to affective and cognitive processes, but not nociceptive pain* (p. 2023.09.21.558825). bioRxiv. https://doi.org/10.1101/2023.09.21.558825

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.

Campbell, P., Foster, N. E., Thomas, E., & Dunn, K. M. (2013). Prognostic Indicators of Low Back Pain in Primary Care: Five-Year Prospective Study. *The Journal of Pain*, *14*(8), 873–883. https://doi.org/10.1016/j.jpain.2013.03.013

Coghill, R. C. (2020). The distributed nociceptive system: A framework for understanding pain. *Trends in Neurosciences*, *43*(10), 780–794. https://doi.org/10.1016/j.tins.2020.07.004

Coghill, R. C., McHaffie, J. G., & Yen, Y.-F. (2003). Neural correlates of interindividual differences in the subjective experience of pain. *Proceedings of the National Academy of Sciences*, *100*(14), 8538–8542.

Cui, Z., & Gong, G. (2018). The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features. *NeuroImage*, *178*, 622–637. https://doi.org/10.1016/j.neuroimage.2018.06.001

Davis, K. D., Aghaeepour, N., Ahn, A. H., Angst, M. S., Borsook, D., Brenton, A., Burczynski, M. E., Crean, C., Edwards, R., Gaudilliere, B., Hergenroeder, G. W.,

Iadarola, M. J., Iyengar, S., Jiang, Y., Kong, J.-T., Mackey, S., Saab, C. Y., Sang, C. N., Scholz, J., … Pelleymounter, M. A. (2020). Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: Challenges and opportunities. *Nature Reviews Neurology*, *16*(7), Article 7. https://doi.org/10.1038/s41582-020-0362-2

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Ding, Y.-S., Singhal, T., Planeta-Wilson, B., Gallezot, J.-D., Nabulsi, N., Labaree, D., Ropchan, J., Henry, S., Williams, W., Carson, R. E., Neumeister, A., & Malison, R. T. (2010). PET imaging of the effects of age and cocaine on the norepinephrine transporter in the human brain using (S,S)-[11C]O-methylreboxetine and HRRT. *Synapse*, *64*(1), 30–38. https://doi.org/10.1002/syn.20696

Dukart, J., Holiga, Š., Chatham, C., Hawkins, P., Forsyth, A., McMillan, R., Myers, J., Lingford-Hughes, A. R., Nutt, D. J., Merlo-Pich, E., Risterucci, C., Boak, L., Umbricht, D., Schobel, S., Liu, T., Mehta, M. A., Zelaya, F. O., Williams, S. C., Brown, G., … Sambataro, F. (2018). Cerebral blood flow predicts differential neurotransmitter activity. *Scientific Reports*, *8*(1), 4074. https://doi.org/10.1038/s41598-018-22444-0

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), Article 1. https://doi.org/10.1038/s41592-018-0235-4

Fayaz, A., Croft, P., Langford, R. M., Donaldson, L. J., & Jones, G. T. (2016). Prevalence of chronic pain in the UK: A systematic review and meta-analysis of population studies. *BMJ Open*, *6*(6), e010364. https://doi.org/10.1136/bmjopen-2015-010364

Finn, E. S. (2021). Is it time to put rest to rest? *Trends in Cognitive Sciences*, *25*(12), 1021–1032. https://doi.org/10.1016/j.tics.2021.09.005

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement‐Related effects in fMRI time‐series. *Magnetic Resonance in Medicine*, *35*(3), 346–355. https://doi.org/10.1002/mrm.1910350312

Gallezot, J.-D., Planeta, B., Nabulsi, N., Palumbo, D., Li, X., Liu, J., Rowinski, C., Chidsey, K., Labaree, D., Ropchan, J., Lin, S.-F., Sawant-Basak, A., McCarthy, T. J., Schmidt, A. W., Huang, Y., & Carson, R. E. (2017). Determination of receptor occupancy in the presence of mass dose: [^{11} C]GSK189254 PET imaging of histamine H $_3$ receptor occupancy by PF-03654746. *Journal of Cerebral Blood Flow & Metabolism*, *37*(3), 1095–1107. https://doi.org/10.1177/0271678X16650697

Geuter, S., Boll, S., Eippert, F., & Büchel, C. (2017). Functional dissociation of stimulus intensity encoding and predictive coding of pain in the insula. *eLife*, *6*. https://doi.org/10.7554/eLife.24770

Gim, S., Hong, S.-J., Reynolds Losin, E. A., & Woo, C.-W. (2024). *Temporal Dynamics of Brain Mediation in Predictive Cue-induced Pain Modulation* [Preprint]. Neuroscience. https://doi.org/10.1101/2024.01.23.576786

Gim, S., Lee, D. H., Lee, S., & Woo, C.-W. (2023). *Interindividual differences in pain can be explained by fMRI, sociodemographic, and psychological factors* (p. 2023.07.06.547919). bioRxiv. https://doi.org/10.1101/2023.07.06.547919

Goltermann, O., & Büchel, C. (2023). *Improving fMRI-based prediction of individual pain ratings by accounting for interindividual variance* (p. 2023.07.06.548050). bioRxiv. https://doi.org/10.1101/2023.07.06.548050

Greene, A. S., Gao, S., Scheinost, D., & Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-04920-3

Grouper, H., Eisenberg, E., & Pud, D. (2019). The relationship between sensitivity to pain and conditioned pain modulation in healthy people. *Neuroscience Letters*, *708*, 134333. https://doi.org/10.1016/j.neulet.2019.134333

Henson, R. (2005). What can functional neuroimaging tell the experimental psychologist? *The Quarterly Journal of Experimental Psychology Section A*, *58*(2), 193–233.

Hillmer, A. T., Esterlis, I., Gallezot, J. D., Bois, F., Zheng, M. Q., Nabulsi, N., Lin, S. F., Papke, R. L., Huang, Y., Sabri, O., Carson, R. E., & Cosgrove, K. P. (2016). Imaging of cerebral α4β2* nicotinic acetylcholine receptors with (−)-[18F]Flubatine PET: Implementation of bolus plus constant infusion and sensitivity to acetylcholine in human brain. *NeuroImage*, *141*, 71–80. https://doi.org/10.1016/j.neuroimage.2016.07.026

Hoeppli, M. E., Nahman-Averbuch, H., Hinkle, W. A., Leon, E., Peugh, J., Lopez-Sola, M., King, C. D., Goldschneider, K. R., & Coghill, R. C. (2022). Dissociation between individual differences in self-reported pain intensity and underlying fMRI brain activation. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-31039-3

Horing, B., Sprenger, C., & Büchel, C. (2019). The parietal operculum preferentially encodes heat pain and not salience. *PLOS Biology*, *17*(8), e3000205. https://doi.org/10.1371/journal.pbio.3000205

Hu, L., & Iannetti, G. D. (2019). Neural indicators of perceptual variability of pain across species. *Proceedings of the National Academy of Sciences*, *116*(5), 1782–1791. https://doi.org/10.1073/pnas.1812499116

Johannes, C. B., Le, T. K., Zhou, X., Johnston, J. A., & Dworkin, R. H. (2010). The prevalence of chronic pain in United States adults: Results of an internet-based survey. *The Journal of Pain*, *11*(11), 1230–1239. https://doi.org/10.1016/j.jpain.2010.07.002

Kaller, S., Rullmann, M., Patt, M., Becker, G.-A., Luthardt, J., Girbardt, J., Meyer, P. M., Werner, P., Barthel, H., Bresch, A., Fritz, T. H., Hesse, S., & Sabri, O. (2017). Test–retest measurements of dopamine D1-type receptors using simultaneous PET/MRI imaging. *European Journal of Nuclear Medicine and Molecular Imaging*, *44*(6), 1025–1032. https://doi.org/10.1007/s00259-017-3645-0

Kantonen, T., Karjalainen, T., Isojärvi, J., Nuutila, P., Tuisku, J., Rinne, J., Hietala, J., Kaasinen, V., Kalliokoski, K., Scheinin, H., Hirvonen, J., Vehtari, A., & Nummenmaa, L. (2020). Interindividual variability and lateralization of μ-opioid

receptors in the human brain. *NeuroImage*, *217*, 116922. https://doi.org/10.1016/j.neuroimage.2020.116922

Knecht, S., Kunesch, E., & Schnitzler, A. (1996). Parallel and serial processing of haptic information in man: Effects of parietal lesions on sensorimotor hand function. *Neuropsychologia*, *34*(7), 669–687. https://doi.org/10.1016/0028-3932(95)00148-4

Kohoutová, L., Heo, J., Cha, S., Lee, S., Moon, T., Wager, T. D., & Woo, C.-W. (2020). Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nature Protocols*, *15*(4), Article 4. https://doi.org/10.1038/s41596-019-0289-5

Kotikalapudi, R., Kincses, B., Zunhammer, M., Schlitt, F., Schmidt-Wilcke, T., Kincses, Z. T., Asan, L., Bingel, U., & Spisak, T. (2022). *Gray matter cortical thickness predicts individual pain sensitivity: A multi-center machine learning approach* (p. 2022.06.14.496092). bioRxiv. https://doi.org/10.1101/2022.06.14.496092

Larivière, S., Bayrak, Ş., Vos de Wael, R., Benkarim, O., Herholz, P., Rodriguez-Cruces, R., Paquola, C., Hong, S.-J., Misic, B., Evans, A. C., Valk, S. L., & Bernhardt, B. C. (2023). BrainStat: A toolbox for brain-wide statistics and multimodal feature associations. *NeuroImage*, *266*, 119807. https://doi.org/10.1016/j.neuroimage.2022.119807

Lee, D. H., Lee, S., & Woo, C.-W. (2023). *Decoding Pain: Uncovering the Factors that Affect Performance of Neuroimaging-Based Pain Models* (p. 2023.12.22.573021). bioRxiv. https://doi.org/10.1101/2023.12.22.573021

Lee, J.-J., Kim, H. J., Čeko, M., Park, B., Lee, S. A., Park, H., Roy, M., Kim, S.-G., Wager, T. D., & Woo, C.-W. (2021). A neuroimaging biomarker for sustained experimental and clinical pain. *Nature Medicine*, 1–9. https://doi.org/10.1038/s41591-020-1142-7

Liang, M., Su, Q., Mouraux, A., & Iannetti, G. D. (2019). Spatial patterns of brain activity preferentially reflecting transient pain and stimulus intensity. *Cerebral Cortex*, *29*(5), 2211–2227. https://doi.org/10.1093/cercor/bhz026

Liu, S., Abdellaoui, A., Verweij, K. J. H., & van Wingen, G. A. (2023). Replicable brain–phenotype associations require large-scale neuroimaging data. *Nature Human Behaviour*, 1–13. https://doi.org/10.1038/s41562-023-01642-5

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., … Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, *603*(7902), Article 7902. https://doi.org/10.1038/s41586-022-04492-9

Markello, R. D., Hansen, J. Y., Liu, Z.-Q., Bazinet, V., Shafiei, G., Suárez, L. E., Blostein, N., Seidlitz, J., Baillet, S., Satterthwaite, T. D., Chakravarty, M. M., Raznahan, A., & Misic, B. (2022). neuromaps: Structural and functional interpretation of brain maps. *Nature Methods*, *19*(11), Article 11. https://doi.org/10.1038/s41592-022-01625-w

Melum, T. A., Vangberg, T. R., Johnsen, L.-H., Steingrímsdóttir, Ó. A., Stubhaug, A., Mathiesen, E. B., & Nielsen, C. (2023). Gray matter volume and pain tolerance in

a general population: The Tromsø study. *PAIN*, *164*(8), 1750. https://doi.org/10.1097/j.pain.0000000000002871

Mogil, J. S. (2021). Sources of individual differences in pain. *Annual Review of Neuroscience*, *44*(1), 1–25. https://doi.org/10.1146/annurev-neuro-092820-105941

Mouraux, A., Diukova, A., Lee, M. C., Wise, R. G., & Iannetti, G. D. (2011). A multisensory investigation of the functional significance of the "pain matrix." *NeuroImage*, *54*(3), 2237–2249.

Mouraux, A., & Iannetti, G. D. (2018). The search for pain biomarkers in the human brain. *Brain*, *141*(12), 3290–3307. https://doi.org/10.1093/brain/awy281

Naganawa, M., Nabulsi, N., Henry, S., Matuskey, D., Lin, S.-F., Slieker, L., Schwarz, A. J., Kant, N., Jesudason, C., Ruley, K., Navarro, A., Gao, H., Ropchan, J., Labaree, D., Carson, R. E., & Huang, Y. (2021). First-in-Human Assessment of 11C-LSN3172176, an M1 Muscarinic Acetylcholine Receptor PET Radiotracer. *Journal of Nuclear Medicine*, *62*(4), 553–560. https://doi.org/10.2967/jnumed.120.246967

Nielsen, C. S., Staud, R., & Price, D. D. (2009). Individual differences in pain sensitivity: Measurement, causation, and consequences. *The Journal of Pain*, *10*(3), 231–237. https://doi.org/10.1016/j.jpain.2008.09.010

Nørgaard, M., Beliveau, V., Ganz, M., Svarer, C., Pinborg, L. H., Keller, S. H., Jensen, P. S., Greve, D. N., & Knudsen, G. M. (2021). A high-resolution in vivo atlas of the human brain's benzodiazepine binding site of GABAA receptors. *NeuroImage*, *232*, 117878.

Normandin, M. D., Zheng, M.-Q., Lin, K.-S., Mason, N. S., Lin, S.-F., Ropchan, J., Labaree, D., Henry, S., Williams, W. A., Carson, R. E., Neumeister, A., & Huang, Y. (2015). Imaging the Cannabinoid CB1 Receptor in Humans with [^{11} C] OMAR: Assessment of Kinetic Analysis Methods, Test–Retest Reproducibility, and Gender Differences. *Journal of Cerebral Blood Flow & Metabolism*, *35*(8), 1313–1322. https://doi.org/10.1038/jcbfm.2015.46

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), Article 2. https://doi.org/10.1038/nrn.2016.167

Radhakrishnan, R., Nabulsi, N., Gaiser, E., Gallezot, J.-D., Henry, S., Planeta, B., Lin, S., Ropchan, J., Williams, W., Morris, E., D'Souza, D. C., Huang, Y., Carson, R. E., & Matuskey, D. (2018). Age-Related Change in 5-HT6 Receptor Availability in Healthy Male Volunteers Measured with 11C-GSK215083 PET. *Journal of Nuclear Medicine*, *59*(9), 1445–1450. https://doi.org/10.2967/jnumed.117.206516

Ronga, I., Valentini, E., Mouraux, A., & Iannetti, G. D. (2012). Novelty is not enough: Laser-evoked potentials are determined by stimulus saliency, not absolute novelty. *Journal of Neurophysiology*, *109*(3), 692–701. https://doi.org/10.1152/jn.00464.2012

Sandiego, C. M., Gallezot, J.-D., Lim, K., Ropchan, J., Lin, S., Gao, H., Morris, E. D., & Cosgrove, K. P. (2015). Reference Region Modeling Approaches for Amphetamine

Challenge Studies with [^{11} C]FLB 457 and PET. *Journal of Cerebral Blood Flow & Metabolism*, *35*(4), 623–629. https://doi.org/10.1038/jcbfm.2014.237

Savli, M., Bauer, A., Mitterhauser, M., Ding, Y.-S., Hahn, A., Kroll, T., Neumeister, A., Haeusler, D., Ungersboeck, J., Henry, S., Isfahani, S. A., Rattay, F., Wadsak, W., Kasper, S., & Lanzenberger, R. (2012). Normative database of the serotonergic system in healthy subjects using multi-tracer PET. *NeuroImage*, *63*(1), 447–459. https://doi.org/10.1016/j.neuroimage.2012.07.001

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Schulz, E., Zherdin, A., Tiemann, L., Plant, C., & Ploner, M. (2012). Decoding an individual's sensitivity to pain from the multivariate analysis of EEG data. *Cerebral Cortex*, *22*(5), 1118–1123. https://doi.org/10.1093/cercor/bhr186

Schulz, M.-A., Bzdok, D., Haufe, S., Haynes, J.-D., & Ritter, K. (2024). Performance reserves in brain-imaging-based phenotype prediction. *Cell Reports*, *43*(1), 113597. https://doi.org/10.1016/j.celrep.2023.113597

Schulz, M.-A., Yeo, B. T. T., Vogelstein, J. T., Mourao-Miranada, J., Kather, J. N., Kording, K., Richards, B., & Bzdok, D. (2020). Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-18037-z

Skovbjerg, S., Jørgensen, T., Arendt-Nielsen, L., Ebstrup, J. F., Carstensen, T., & Graven-Nielsen, T. (2017). Conditioned Pain Modulation and Pressure Pain Sensitivity in the Adult Danish General Population: The DanFunD Study. *The Journal of Pain*, *18*(3), 274–284. https://doi.org/10.1016/j.jpain.2016.10.022

Smart, K., Cox, S. M. L., Scala, S. G., Tippler, M., Jaworska, N., Boivin, M., Séguin, J. R., Benkelfat, C., & Leyton, M. (2019). Sex differences in [11C]ABP688 binding: A positron emission tomography study of mGlu5 receptors. *European Journal of Nuclear Medicine and Molecular Imaging*, *46*(5), 1179–1183. https://doi.org/10.1007/s00259-018-4252-4

Spisak, T., Bingel, U., & Wager, T. D. (2023). Multivariate BWAS can be replicable with moderate sample sizes. *Nature*, *615*(7951), Article 7951. https://doi.org/10.1038/s41586-023-05745-x

Spisak, T., Kincses, B., Schlitt, F., Zunhammer, M., Schmidt-Wilcke, T., Kincses, Z. T., & Bingel, U. (2020). Pain-free resting-state functional brain connectivity predicts individual pain sensitivity. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-019-13785-z

Starr, C. J., Sawaki, L., Wittenberg, G. F., Burdette, J. H., Oshiro, Y., Quevedo, A. S., & Coghill, R. C. (2009). Roles of the insular cortex in the modulation of pain: Insights from brain lesions. *Journal of Neuroscience*, *29*(9), 2684–2694. https://doi.org/10.1523/JNEUROSCI.5173-08.2009

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245–251. https://doi.org/10.1037/0033-2909.87.2.245

Su, Q., Qin, W., Yang, Q. Q., Yu, C. S., Qian, T. Y., Mouraux, A., Iannetti, G. D., & Liang, M. (2019). Brain regions preferentially responding to transient and iso-intense

painful or tactile stimuli. *NeuroImage*, *192*, 52–65. https://doi.org/10.1016/j.neuroimage.2019.01.039

Su, Q., Song, Y., Zhao, R., & Liang, M. (2020). A review on the ongoing quest for a pain signature in the human brain. *Brain Science Advances*, *5*(4), 274–287.

Szucs, D., & Ioannidis, J. PA. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage*, *221*, 117164. https://doi.org/10.1016/j.neuroimage.2020.117164

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tracey, I. (2021). Neuroimaging enters the pain biomarker arena. *Science Translational Medicine*, *13*(619), eabj7358. https://doi.org/10.1126/scitranslmed.abj7358

Tu, Y., Li, Z., Zhang, L., Zhang, H., Bi, Y., Yue, L., & Hu, L. (2023). Pain-preferential thalamocortical neural dynamics across species. *Nature Human Behaviour*, 1–15. https://doi.org/10.1038/s41562-023-01714-6

Tu, Y., Zhang, B., Cao, J., Wilson, G., Zhang, Z., & Kong, J. (2019). Identifying inter-individual differences in pain threshold using brain connectome: A test-retest reproducible study. *NeuroImage*, *202*, 116049. https://doi.org/10.1016/j.neuroimage.2019.116049

Turner, B. O., Paul, E. J., Miller, M. B., & Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Communications Biology*, *1*(1), Article 1. https://doi.org/10.1038/s42003-018-0073-z

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

Vigotsky, A. D., Jabakhanji, R., Branco, P., Iannetti, G. D., Baliki, M. N., & Apkarian, A. V. (2022). *Widespread, perception-related information in the human brain scales with levels of consciousness* (p. 2022.09.19.508437). bioRxiv. https://doi.org/10.1101/2022.09.19.508437

Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C.-W., & Kross, E. (2013). An fMRI-based neurologic signature of physical pain. *New England Journal of Medicine*, *368*(15), 1388–1397.

Wei, Z., Duan, Y., Zhu, Y., Lin, X., Zhang, M., Brooks, J. C. W., Liu, Y., Hu, L., & Kong, Y. (2024). *Cortico-spinal Mechanisms of Periphery Neuromodulation induced Analgesia* (p. 2024.02.06.579059). bioRxiv. https://doi.org/10.1101/2024.02.06.579059

Wen, Z., Raio, C. M., Pace-Schott, E. F., Lazar, S. W., LeDoux, J. E., Phelps, E. A., & Milad, M. R. (2022). Temporally and anatomically specific contributions of the human amygdala to threat and safety learning. *Proceedings of the National Academy of Sciences*, *119*(26), e2204066119. https://doi.org/10.1073/pnas.2204066119

Werner, M. U., Mjöbo, H. N., Nielsen, P. R., Rudin, Å., & Warner, D. S. (2010). Prediction of Postoperative Pain: A Systematic Review of Predictive Experimental Pain

Studies. *Anesthesiology*, *112*(6), 1494–1502. https://doi.org/10.1097/ALN.0b013e3181dcd5a0

Woo, C.-W., Schmidt, L., Krishnan, A., Jepma, M., Roy, M., Lindquist, M. A., Atlas, L. Y., & Wager, T. D. (2017). Quantifying cerebral contributions to pain beyond nociception. *Nature Communications*, *8*(1), 1–14. https://doi.org/10.1038/ncomms14211

Woolf, C. J., & Max, M. B. (2001). Mechanism-based Pain Diagnosis: Issues for Analgesic Drug Development. *Anesthesiology*, *95*(1), 241–249. https://doi.org/10.1097/00000542-200107000-00034

Xu, A., Larsen, B., Baller, E. B., Scott, J. C., Sharma, V., Adebimpe, A., Basbaum, A. I., Dworkin, R. H., Edwards, R. R., Woolf, C. J., Eickhoff, S. B., Eickhoff, C. R., & Satterthwaite, T. D. (2020). Convergent neural representations of experimentally-induced acute pain in healthy volunteers: A large-scale fMRI meta-analysis. *Neuroscience & Biobehavioral Reviews*, *112*, 300–323. https://doi.org/10.1016/j.neubiorev.2020.01.004

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670.

Yeung, A. W. K. (2018). An Updated Survey on Statistical Thresholding and Sample Size of fMRI Studies. *Frontiers in Human Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnhum.2018.00016

Zhang, L.-B., Lu, X.-J., Huang, G., Zhang, H.-J., Tu, Y.-H., Kong, Y.-Z., & Hu, L. (2022). Selective and replicable neuroimaging-based indicators of pain discriminability. *Cell Reports Medicine*, *3*(12), 100846. https://doi.org/10.1016/j.xcrm.2022.100846

Zhang, X., Li, L., Huang, G., Zhang, L., Liang, Z., Shi, L., & Zhang, Z. (2021). A multisensory fMRI investigation of nociceptive-preferential cortical regions and responses. *Frontiers in Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fnins.2021.635733

Zubieta, J.-K., Smith, Y. R., Bueller, J. A., Xu, Y., Kilbourn, M. R., Jewett, D. M., Meyer, C. R., Koeppe, R. A., & Stohler, C. S. (2001). Regional Mu Opioid Receptor Regulation of Sensory and Affective Dimensions of Pain. *Science*, *293*(5528), 311–315. https://doi.org/10.1126/science.1060952

Figure legends

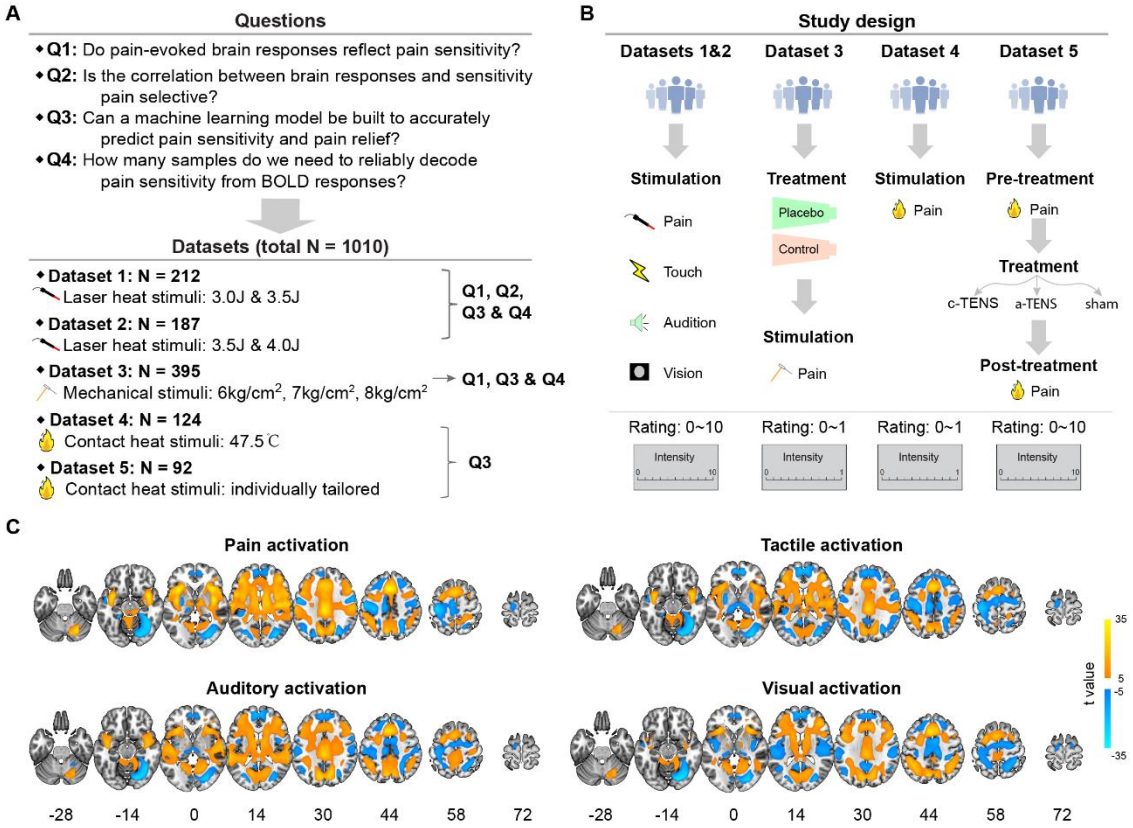

Figure 1. Study overview and brain activations responsive to pain and nonpain stimuli. (A) Key questions and datasets used to answer these questions. Using five large datasets with laser heat pain, mechanical pain, and contact heat pain stimuli of different physical intensity, we examined the correlation between pain-evoked BOLD responses and pain sensitivity, tested pain selectivity of this correlation, built machine learning models to predict pain sensitivity and pain relief, and explored sample sizes needed to reliably reveal the correlation between BOLD responses and pain sensitivity. **(B)** Study design of five datasets. In all five datasets, participants received sensory stimuli and rated the perceived intensity with a rating scale (from 0 to 10 or from 0 to1). In Datasets 3&5, participants also received treatments (placebo in Dataset 4 and transcutaneous electrical nerve stimulation [TENS]) to relive pain. **(C)** Brain areas activated by pain and nonpain stimuli in Datasets 1&2. Painful, tactile, auditory, and visual stimuli activated largely overlapped brain areas. c-TENS: conventional TENS; a-TENS: acupuncture-like TENS.

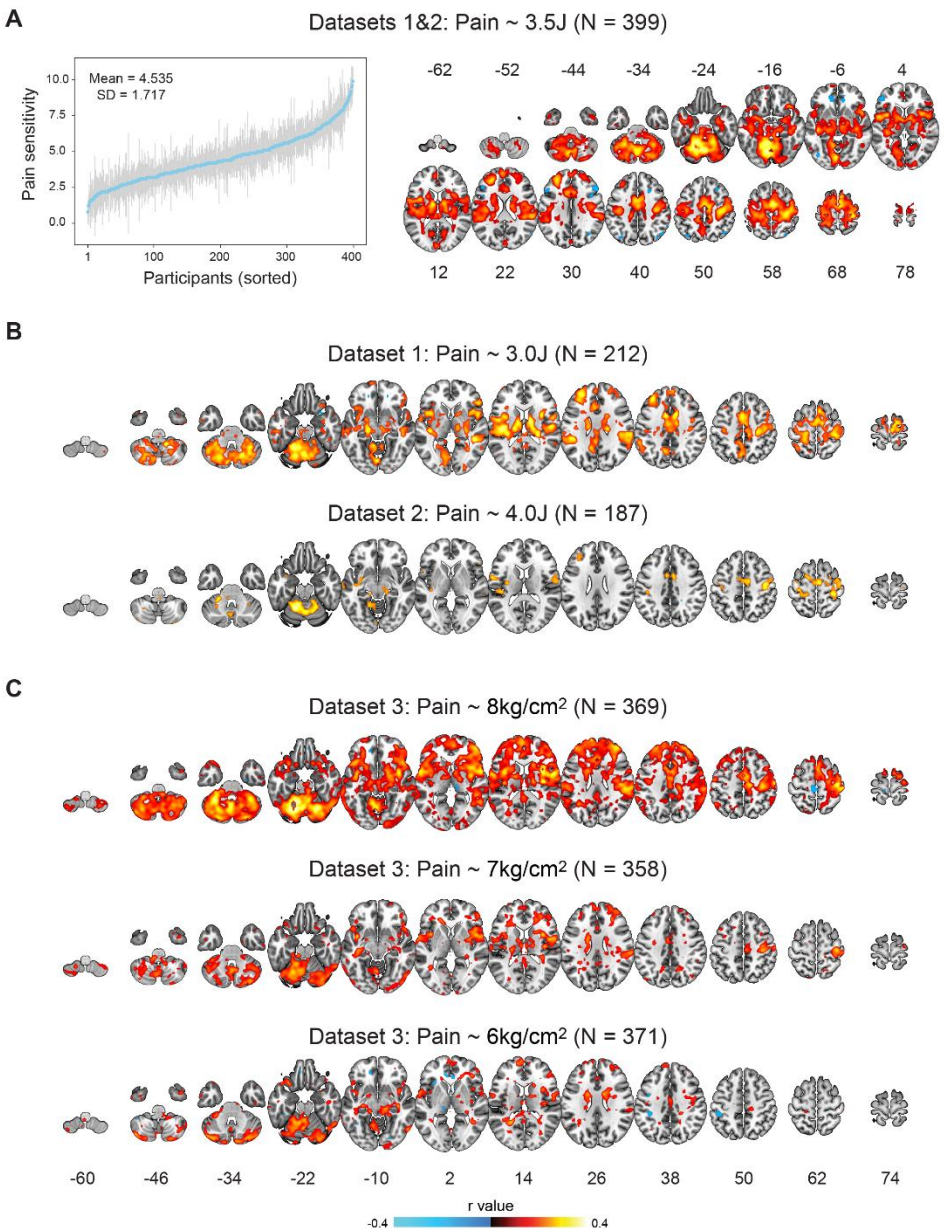

Figure 2. Correlations between pain-evoked BOLD responses and pain sensitivity. (A) BOLD-pain sensitivity correlations in the 3.5J condition in Datasets 1&2. Pain sensitivity varied considerably across individuals. This interindividual variability reflected in pain-evoked BOLD responses in a wide range of areas. **(B)** Replication of the correlation in other conditions in Datasets 1&2. BOLD responses reliably correlated with pain sensitivity in the 3.0J and 4.0J conditions. **(C)** Replication of the correlation in Dataset 3. BOLD responses correlated with mechanical pain sensitivity in all conditions, demonstrating the replicability and generalizability of the correlation. Error bars in (A) stand for standard

deviation of intensity ratings for each participant. Please note that the correlation between BOLD responses and pain sensitivity in Dataset 4 was previously reported, and similar correlational analyses could not be conducted in Dataset 5 since the stimulus intensity differed for different participants.

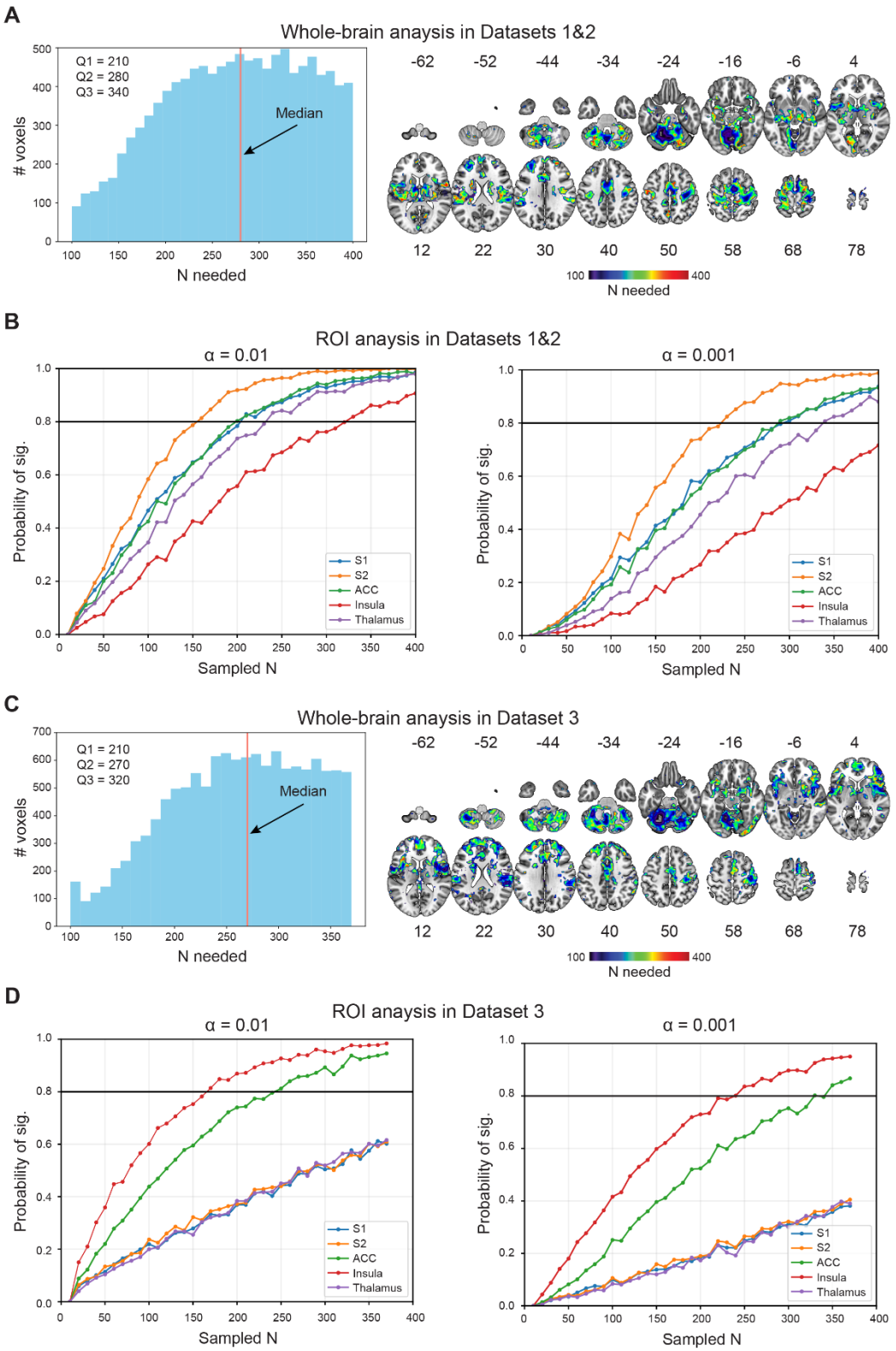

Figure 3. Influence of sample sizes on the detectability of the BOLD-pain sensitivity correlation. (A) Sample sizes needed to detect correlation with a probability ≥ 0.8 in whole

brain analysis in Datasets 1&2. Different sample sizes were needed to reliably detect the correlations in different voxels, and only half the voxels required a sample size less than 280. **(B)** Sample sizes needed to detect correlation in ROI analysis in Datasets 1&2. In ROI analysis, large sample sizes were still needed to reliably detect significant correlation. **(C)** Sample sizes needed to detect correlation with a probability ≥ 0.8 in whole brain analysis in Dataset 3. **(D)** Sample sizes needed to detect correlation in ROI analysis in Dataset 3.

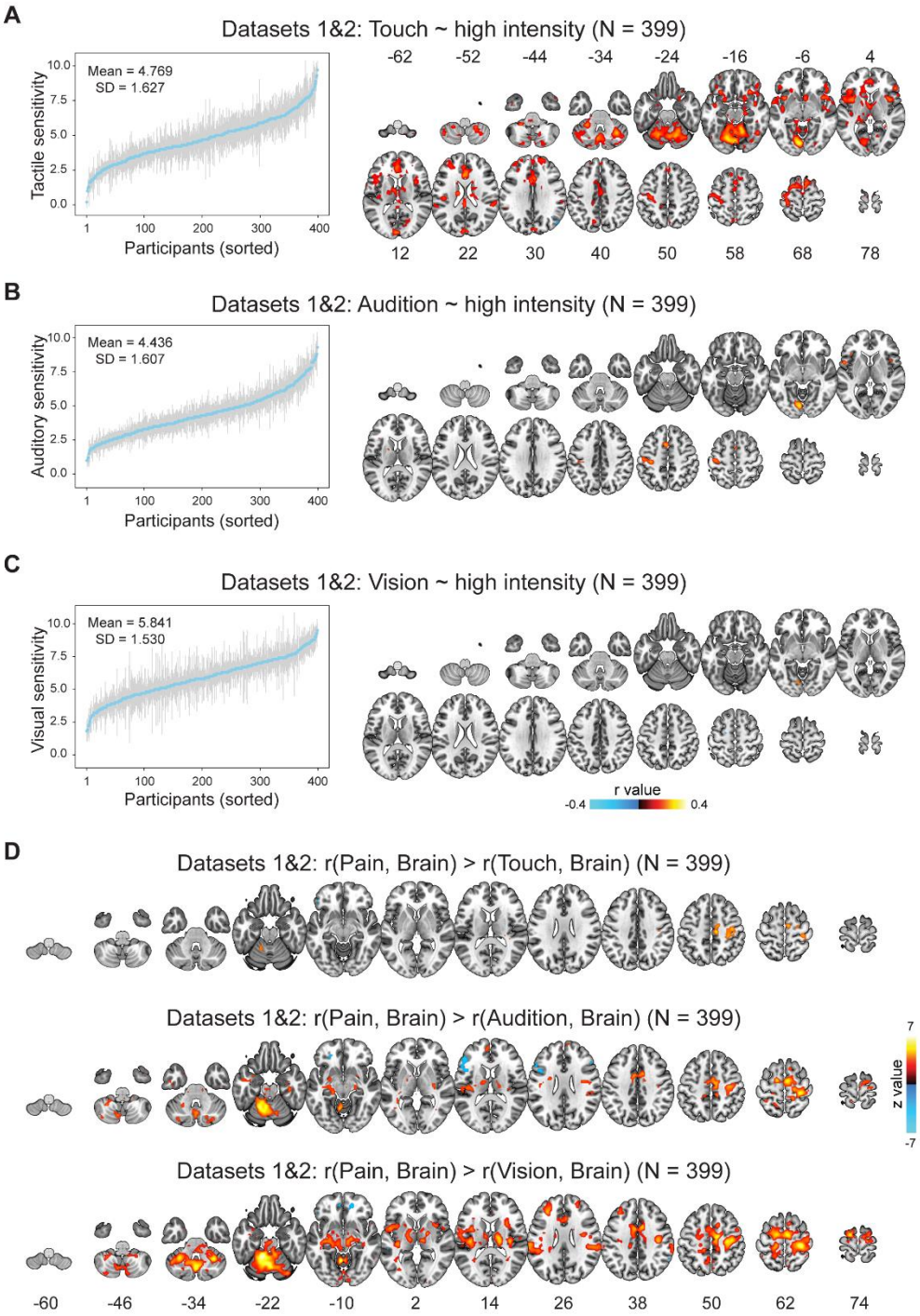

Figure 4. Correlations between BOLD responses and sensory sensitivity in nonpain modalities. (A)~(C) BOLD-sensitivity correlation in the high intensity condition in Datasets 1&2 for touch (A), audition (B), and vision (C). Interindividual variability of tactile sensitivity correlated with BOLD responses in many areas, while only a few areas showed correlation with auditory and visual sensitivity. **(D)** Comparisons of BOLD-

sensitivity correlation between pain and nonpain modalities. Direct comparisons of correlation coefficients showed larger correlations for pain than for touch, audition, and vision. Error bars in (A)~(C) stand for standard deviation of intensity ratings for each participant.

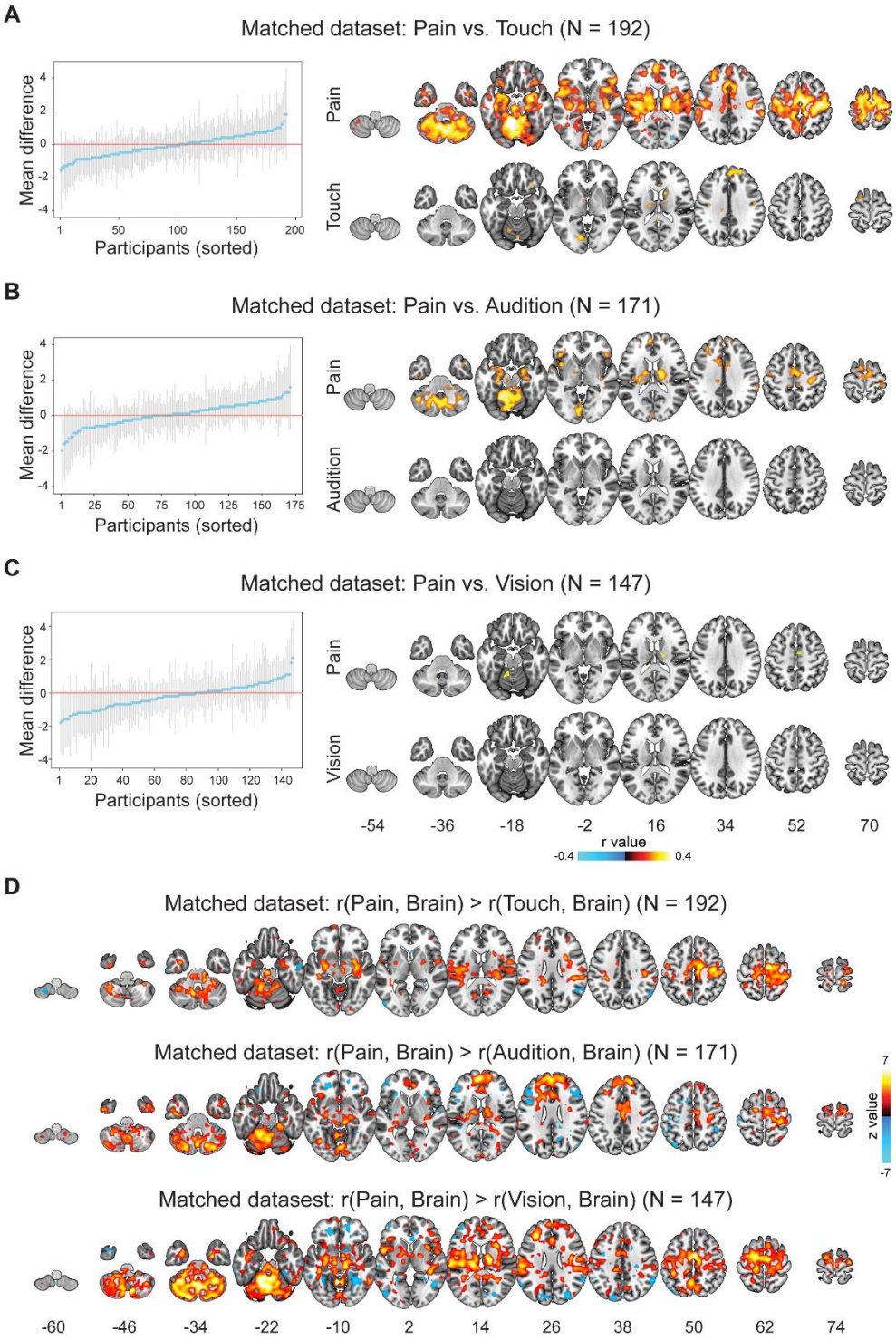

Figure 5. Correlation between BOLD responses and sensory sensitivity in nonpain modalities after sensitivity matching. (A)~(C) BOLD-sensitivity correlation for pain and nonpain after sensitivity matching. In sensitivity matching, we selected participants whose mean pain and tactile (A), auditory (B), or visual (C) ratings did not differ significantly.

After the pain-touch and pain-audition matching, pain-evoked BOLD responses still correlated with pain sensitivity. After pain-vision matching, only a few areas showed correlation with pain sensitivity. In contrast, far less areas correlated with tactile/auditory/visual sensitivity in all matchings. **(D)** Comparisons of BOLD-sensitivity correlation between pain and nonpain modalities after sensitivity matching. Direct comparisons of correlation coefficients showed larger correlations for pain than for touch, audition, and vision. Error bars in (A)~(C) stand for 95% confidence intervals for the rating differences for each participant.

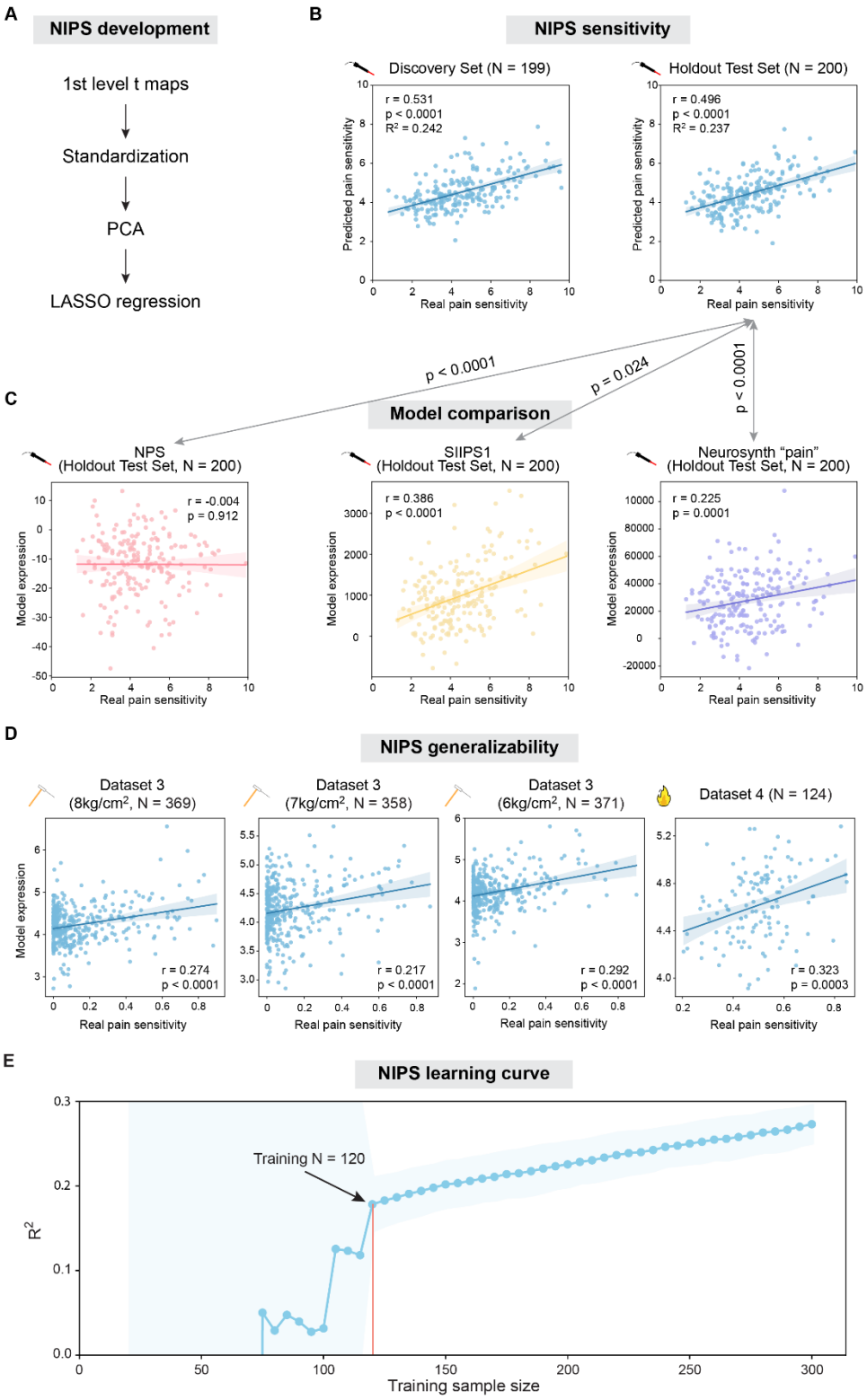

Figure 6. Development of NIPS and its performance. (A) Development of NIPS. We used 1st t maps for pain as features and built our model (NIPS) with LASSO-PCR. **(B)** Performance of NIPS in the Discovery Set and Holdout Test Set. In both datasets, NIPS

showed good performance, explaining about 20% of the variance of pain sensitivity. **(C)** Comparisons of NIPS and other models. Some well-established models performed not as good as NIPS, which was specifically developed to predict pain sensitivity with laser heat pain data. **(D)** Generalizability of NIPS. Although developed with laser heat pain data, NIPS significantly predicted mechanical and contact heat pain sensitivity in external datasets. **(E)** Learning curve of NIPS. NIPS's performance varied dramatically with small training sample sizes, but stabilized and continuously improved as training sample sizes exceeded 120. Shaded regions in (B)~(D) stand for 95% confidence intervals, but shaded regions in (E) represent standard deviation of R^2.

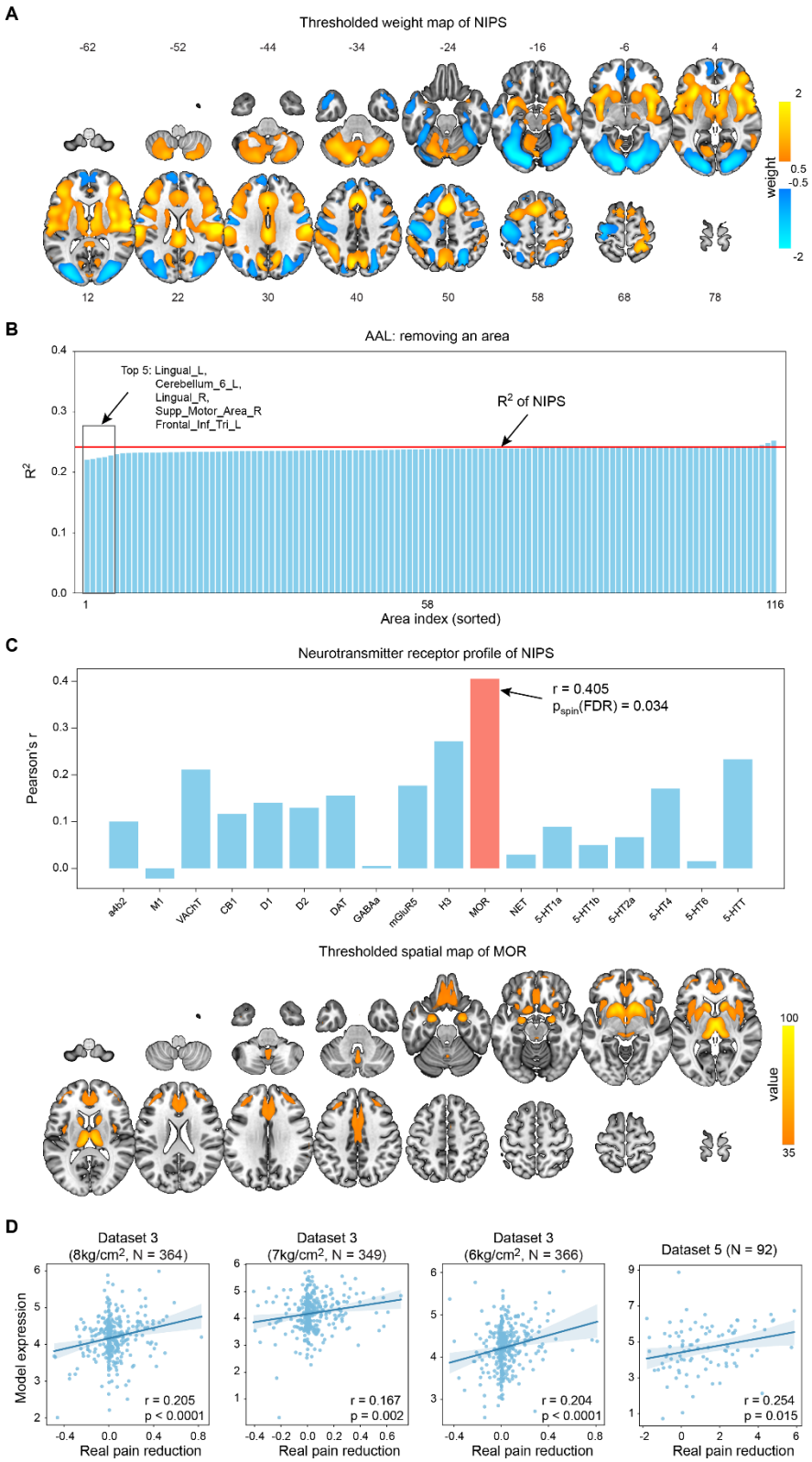

Figure 7. Interpretation of the weight map of NIPS. (A) Thresholded weight map of NIPS. For the sake of visualization, the weight map was thresholded at 0.5/-0.5 (p(FDR) <

0.05 for all surviving voxels). Model prediction was made with the unthresholded map. **(B)** Virtual lesion of NIPS. Removing brain regions one by one did not make NIPS's performance substantially worse, suggesting that a distributed network rather than isolated areas is responsible for encoding pain sensitivity. **(C)** Neurotransmitter profile of NIPS and thresholded spatial distribution of MOR. Assessed by the spin test, the weight map of NIPS correlated significantly with the spatial map of MOR. **(D)** Prediction of pain relief by NIPS. In Datasets 3&5, NIPS significantly predicted pain reduction after pain treatments. a4b2: alpha-4 beta-2 nicotinic receptor; M1: muscarinic acetylcholine receptor subtype M1; VAChT: vesicular acetylcholine transporter; CB1: cannabinoid type-1 receptor; D1: dopamine D1-type receptor; D2: dopamine D2-type receptor; DAT: respective dopamine transporter; mGluR5: metabotropic glutamate type 5 receptor; H3: histamine H3 receptors; MOR: μ-opioid receptor; NET: norepinephrine transporter; 5-HT1a: serotonin 1A receptor (18) 5-HTT: serotonin transporter. Shaded regions in (D) stand for 95% confidence intervals.