

ORIGINAL ARTICLE

An fMRI-Based Neurologic Signature of Physical Pain

Tor D. Wager, Ph.D., Lauren Y. Atlas, Ph.D., Martin A. Lindquist, Ph.D., Mathieu Roy, Ph.D., Choong-Wan Woo, M.A., and Ethan Kross, Ph.D.

ABSTRACT

BACKGROUND

From the Department of Psychology and Neuroscience, University of Colorado, Boulder (T.D.W., M.R., C.-W.W.); the Department of Psychology, New York University, New York (L.Y.A.); the Department of Biostatistics, Johns Hopkins University, Baltimore (M.A.L.); and the Department of Psychology, University of Michigan, Ann Arbor (E.K.). Address reprint requests to Dr. Wager at the Department of Psychology and Neuroscience, University of Colorado, Boulder, 345 UCB, Boulder, CO 80305, or at tor.wager@colorado.edu.

N Engl J Med 2013;368:1388-97.
DOI: 10.1056/NEJMoa1204471
Copyright © 2013 Massachusetts Medical Society.

METHODS

In four studies involving a total of 114 participants, we developed an fMRI-based measure that predicts pain intensity at the level of the individual person. In study 1, we used machine-learning analyses to identify a pattern of fMRI activity across brain regions — a neurologic signature — that was associated with heat-induced pain. The pattern included the thalamus, the posterior and anterior insulae, the secondary somatosensory cortex, the anterior cingulate cortex, the periaqueductal gray matter, and other regions. In study 2, we tested the sensitivity and specificity of the signature to pain versus warmth in a new sample. In study 3, we assessed specificity relative to social pain, which activates many of the same brain regions as physical pain. In study 4, we assessed the responsiveness of the measure to the analgesic agent remifentanil.

RESULTS

In study 1, the neurologic signature showed sensitivity and specificity of 94% or more (95% confidence interval [CI], 89 to 98) in discriminating painful heat from nonpainful warmth, pain anticipation, and pain recall. In study 2, the signature discriminated between painful heat and nonpainful warmth with 93% sensitivity and specificity (95% CI, 84 to 100). In study 3, it discriminated between physical pain and social pain with 85% sensitivity (95% CI, 76 to 94) and 73% specificity (95% CI, 61 to 84) and with 95% sensitivity and specificity in a forced-choice test of which of two conditions was more painful. In study 4, the strength of the signature response was substantially reduced when remifentanil was administered.

CONCLUSIONS

It is possible to use fMRI to assess pain elicited by noxious heat in healthy persons. Future studies are needed to assess whether the signature predicts clinical pain. (Funded by the National Institute on Drug Abuse and others.)

ALTHOUGH BIOMARKERS FOR MEDICAL conditions have proliferated over the past 50 years, objective assessments related to mental health have lagged behind. Physical pain is an affliction associated with enormous cognitive, social, and economic costs,¹ but pain is not easy to ascertain. It is primarily assessed by means of self-report, an imperfect measure of subjective experience. The capacity to effectively report pain is limited in many vulnerable populations (e.g., the very old or very young, persons with cognitive impairment, and those who are minimally conscious). Moreover, self-report provides a limited basis for understanding the neurophysiological processes underlying different types of pain and thus a limited basis for targeting treatments to the underlying neuropathologic conditions. As a result, current approaches to pain assessment focus on a convergence of biologic, behavioral, and self-report measures.²

It is plausible that neurologic signatures (patterns of activity across brain regions) derived from brain imaging could provide direct measures of pain intensity and be used to compare analgesic treatments.³ We combined the use of functional magnetic resonance imaging (fMRI) with machine learning^{4,5} to develop a brain-based neurologic signature for experimental thermal pain.

METHODS

PARTICIPANTS

The studies included a total of 114 healthy participants. Study 1 included 20 participants, 8 of whom were women; the mean (\pm SD) age was 28.8 ± 7.5 years. Study 2 included 33 participants, 22 of whom were women; the mean age was 27.9 ± 9.0 years. Study 3 included 40 participants, 21 of whom were women; the mean age was 20.8 ± 2.6 years.⁶ Study 4 included 21 participants, 11 of whom were women; the mean age was 24.7 ± 4.2 years.⁷ The Columbia University institutional review board approved all the studies, and all participants provided written informed consent. All the authors vouch for the accuracy and completeness of the data and analyses reported and the fidelity of the studies to the protocols. See the Supplementary Appendix, available with the full text of this article at NEJM.org, for additional details.

STUDY DESIGN

In all four studies, we applied thermal stimuli in randomized sequences of varying intensity (trials) to the left forearm of each participant during fMRI scanning. For imaging, we used a 1.5-T General Electric scanner in studies 1, 3, and 4 and a 3-T Phillips scanner in study 2.

Participants in study 1 underwent 12 trials at each of four intensities, which were calibrated for each person: innocuous warmth (defined with the use of self-report by the participant as level 1 on a 9-point visual-analogue scale [VAS], with a mean [\pm SD] temperature of $41.0 \pm 1.9^\circ\text{C}$) and three levels of painful heat (participant-defined levels 3, 5, and 7, with mean temperatures of $43.3 \pm 2.1^\circ\text{C}$, $45.4 \pm 1.71^\circ\text{C}$, and $47.1 \pm 0.98^\circ\text{C}$, respectively). Each trial consisted of a warning cue and anticipation period (8 seconds), stimulation (10 seconds), and a pain-recall and rating period (4 seconds), with periods of rest before and after recall.

Participants in study 2 underwent a total of 75 trials across six temperatures (44.3 to 49.3°C in 1°C increments). After each trial, participants judged whether the stimulus was painful. They subsequently judged nonpainful warmth on a 100-point VAS and pain intensity on a 100-point VAS. Ratings were coded from 0 to 99 for nonpainful events and from 100 to 200 for painful events.

Participants in study 3 underwent 32 trials, consisting of 8 trials with each of four stimulus types. We delivered noxious heat ($46.6 \pm 1.7^\circ\text{C}$, denoted “painful”) and warmth that was near the pain threshold ($39.9 \pm 2.8^\circ\text{C}$, denoted “warm”) at temperatures calibrated for each person. Each participant had recently experienced a romantic breakup and continued to feel intensely rejected. During scanning, participants viewed an image of their ex-partner (denoted as “rejecter” trials, which elicit social pain⁸) and an image of a close friend (denoted as “friend” trials).

Participants in study 4 received two intravenous infusions of remifentanil, a potent μ -opioid agonist, during fMRI scanning in two series of trials. In the open-infusion series, participants knew they received remifentanil, and in the hidden-infusion series, they were told that no drug was delivered, even though it had been administered. Remifentanil doses (mean dose, $0.043 \pm 0.01 \mu\text{g}$ per kilogram of body weight per minute) were individually calibrated before the

session to elicit analgesia without sedation, and we estimated the brain concentration of the drug over time using a pharmacokinetic model.⁹ We conducted 36 trials — 18 involving pain (mean temperature, $47.1 \pm 1.7^\circ\text{C}$) and 18 involving warmth (mean temperature, $41.2 \pm 2.6^\circ\text{C}$) — during each of the two infusion series. Drug infusion began partway through each series, after 6 trials, and ended after 24 trials. This design resulted in a continuously varying concentration of the drug over time during each infusion series.

DERIVING THE SIGNATURE

In study 1, we used a machine-learning-based regression technique, LASSO-PCR (least absolute shrinkage and selection operator-regularized principal components regression),¹⁰ to predict pain reports from the fMRI activity. We selected relevant brain areas *a priori* using the NeuroSynth meta-analytic database¹¹ (see the Supplementary Appendix) and averaged the brain activity for each intensity level within each participant.¹²⁻¹⁴ We used the signal values from the voxels, each of which measured 3 mm^3 , in the *a priori* map to predict continuous pain ratings, using leave-one-participant-out cross-validation⁴ (see the Supplementary Appendix). The result was a spatial pattern of regression weights across brain regions, which was prospectively applied to fMRI activity maps obtained from new participants. Application of the signature to an activity map (e.g., a map obtained during thermal stimulation) yielded a scalar response value, which constituted the predicted pain for that condition.

We used permutation tests to obtain unbiased estimates of accuracy and bootstrap tests to determine which brain areas made reliable contributions to prediction (Fig. 1). Stimulation did not elicit head movement, and head-movement estimates did not predict pain (for a description of head-movement analyses, see the Supplementary Appendix).

PREDICTING PAIN IN AN INDEPENDENT SAMPLE

In study 2, we tested the neurologic signature identified in study 1, with no further model fitting, for the prediction of pain in individual participants, using data from a different scanner. We also estimated activity maps and signature responses for individual trials, which allowed us to use mixed-effects regression models to test the relationship between neurologic signature re-

sponses and intensity judgments during trials involving painful and nonpainful stimuli.

TESTING FOR SPECIFICITY

In study 3, we applied the signature to activation maps that resulted from physical sensation (painful and warm conditions) and from viewing images related to social pain (rejecter and friend conditions).⁶

RESPONSE TO ANALGESIC TREATMENT

In study 4, we tested the effects of stimulus intensity (painful vs. warm), administration of remifentanil (drug concentration), and manner of drug administration (open vs. hidden) on the signature response. For each of the open and hidden trial series, we estimated activation maps for painful stimulation, warm stimulation, and the magnitude of changes in each that followed the *a priori* time course of drug concentration from the pharmacokinetic model (Fig. S5 in the Supplementary Appendix). Because the drug concentration was continuous over time, the binary classification of painful versus warm conditions was based on the averages of the results of three trials before drug administration and three trials performed at the peak drug concentration.

STATISTICAL ANALYSIS

We assessed the sensitivity and specificity of the signature for two kinds of decisions. In one test, the discrimination of pain from no pain, we compared the signature-response value (i.e., the strength of expression of the signature pattern) for one condition with a threshold, with a response over the threshold being classified as a pain response. Receiver-operating-characteristic plots traced the tradeoff of sensitivity and specificity at different thresholds (Fig. 1D), and the threshold that minimized overall classification errors is reported (Table 1).

In forced-choice discrimination, two activation maps from the same participant were compared, and the image with the higher overall signature response (i.e., the stronger expression of the signature pattern) was classified as associated with more pain. Forced-choice tests are particularly suitable for fMRI because they do not compare the signature response with a threshold that is fixed across persons. Therefore, they do not require people to use the pain-reporting scale in the same way, and they do not require

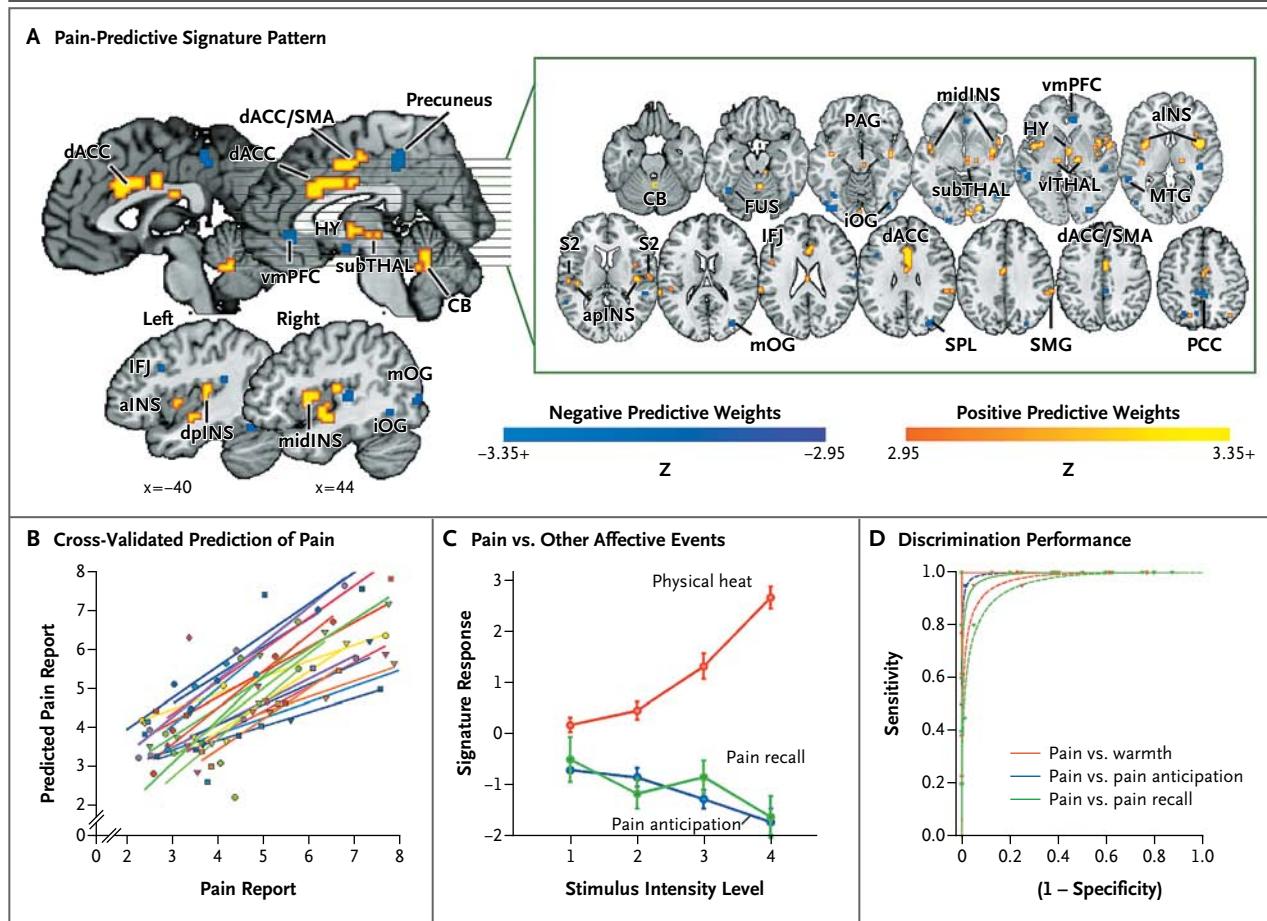


Figure 1. Prediction of Physical Pain on the Basis of Normative Data from Other Participants in Study 1.

Panel A shows the signature map, consisting of voxels in which activity reliably predicted pain. The map shows weights that exceed a threshold (a false discovery rate of $q < 0.05$) for display only; all weights were used in prediction. ACC denotes anterior cingulate cortex, CB cerebellum, FUS fusiform, HY hypothalamus, IFJ inferior frontal junction, INS insula, MTG middle temporal gyrus, OG occipital gyrus, PAG periaqueductal gray matter, PCC posterior cingulate cortex, PFC prefrontal cortex, S2 secondary somatosensory cortex, SMA supplementary motor area, SMG supramarginal gyrus, SPL superior parietal lobe, TG temporal gyrus, and THAL thalamus. Direction is indicated with preceding lowercase letters as follows: a denotes anterior, d dorsal, i inferior, l lateral, m middle, mid mid-insula, p posterior, and v ventral. Panel B shows reported pain versus cross-validated predicted pain. Each colored line or symbol represents an individual participant. Panel C shows the signature response versus the pain intensity for heat, pain-anticipation, and pain-recall conditions. Signature-response values were calculated by taking the dot product of the signature-pattern weights and parameter estimates from a standard, single-participant general linear model, with regressors for each condition. The estimates shown are derived from cross-validation, so that signature weights and test data are independent. I bars indicate standard errors. The receiver-operating-characteristic plots in Panel D show the tradeoff between specificity and sensitivity. Lines are fitted curves, assuming gaussian signal distributions. The test of pain versus no pain and the forced-choice test are shown by dashed lines and solid lines, respectively. Performance on the forced-choice test was at 100% for all conditions; thus, the lines are overlapping.

the scale of fMRI activity to be the same across scanners (see the Supplementary Appendix). Sensitivity, specificity, positive predictive value, and decision accuracy are all equivalent in the forced-choice test. The MATLAB code for implementing all analyses is available at <http://wagerlab.colorado.edu/>.

RESULTS

CROSS-VALIDATED PREDICTION OF PAIN

In study 1, the neurologic signature included significant positive weights in regions including the bilateral dorsal posterior insula, the secondary somatosensory cortex, the anterior insula, the ventro-

Table 1. Pain-Classification Performance, According to Study.*

Study	Discrimination between Pain and No Pain†			Effect Size‡		P Value	Performance on Forced-Choice Test§
	Signature-Response Threshold	Sensitivity	Specificity percent (95% CI)	Positive Predictive Value	AUC		
Study 1							
Painful vs. warm¶	1.40	95 (86–100)	95 (86–100)	0.95	2.69	<0.001	100 (100–100)
Pain vs. pain anticipation	0.36	100 (100–100)	99 (96–100)	0.95	3.69	<0.001	100 (100–100)
Pain vs. pain recall	0.54	95 (85–100)	94 (89–98)	0.96	2.35	<0.001	100 (100–100)
Study 2							
Painful vs. warm **	1.32	93 (84–100)	93 (84–100)	0.92	1.54	<0.001	100 (100–100)
Painful vs. near pain threshold††	2.50	88 (77–97)	85 (72–95)	0.85	1.74	<0.001	100 (100–100)
High vs. low warmth	1.00	56 (36–75)	100 (100–100)	0.79	1.31	0.001	100 (100–100)
Study 3							
Painful vs. warm	1.40‡‡	85 (76–94)	78 (67–89)	0.86	1.64	<0.001	93 (86–98)
Painful vs. rejecter	1.40‡‡	85 (76–94)	73 (61–84)	0.88	1.83	<0.001	95 (89–100)
Rejecter vs. friend	1.40‡‡	27 (16–38)	88 (79–95)	0.57	0.31	0.22	56 (43–69)
Study 4							
Painful vs. warm, before drug treatment	1.40‡‡	90 (79–100)	81 (65–95)	0.83	1.61	<0.001	90 (79–100)
Painful vs. warm, during drug treatment	1.61	86 (73–96)	62 (42–80)	0.74	1.01	0.003	76 (61–90)
Painful before vs. during drug treatment	1.61	86 (72–96)	62 (43–79)	0.74	1.01	0.003	76 (60–92)

* Study 1 included 12 trials each in painful and warm conditions. Study 2 included a mean ($\pm SD$) of 24 \pm 13 trials for pain and 36 \pm 9 trials for warmth, depending on the ratings. Study 3 included 8 trials each in painful and warm conditions. Study 4 included 3 trials for pain and 3 for warmth in the before-drug-treatment condition and in the condition with peak drug concentration. CI denotes confidence interval.

† The tradeoff between sensitivity and specificity at different thresholds was assessed by means of receiver-operating-characteristic (ROC) plots; the signature-response threshold that minimized overall classification errors is reported here.

‡ For the area under the ROC curve, chance is 0.5. Discriminability is a measure of effect size under a gaussian model. Performance varied across studies, according to the number of trials averaged to form the condition maps.

§ For the two-choice (forced-choice) discrimination test, the classification threshold for the difference between paired observations is 0. The sensitivity, specificity, and positive predictive value are the same and are equal to the decision accuracy.

¶ Painful conditions were defined as temperatures greater than 44.5°C and as ratings of more than an average of 5.80 points on a visual-analogue scale (VAS), and warm conditions as temperatures of less than 44.5°C and ratings of less than 3.34 points on the VAS.

|| Study 2 was conducted with the use of a scanner with a different field strength (3 T), so the threshold was reestimated.

** Participants made judgments of painful versus nonpainful conditions for each trial.
†† Participants rated pain or warmth intensity on a continuous VAS, with scores ranging from 0 to 200 points for pain. Pain was defined as a score of more than 125 points, near the pain threshold as a score of 50 to 100 points, high warmth as a score of 0 to 50 points.

‡‡ The threshold derived from study 1 was applied.

lateral and medial thalamus, the hypothalamus, and the dorsal anterior cingulate cortex ($q<0.05$, corrected for the false discovery rate) (Fig. 1A, and Table S1 in the Supplementary Appendix), which is consistent with the view of pain as a distributed process.^{15,16} In a leave-one-participant-out cross-validation test, the neurologic signature accurately predicted continuous pain ratings, with a mean ($\pm SD$) error of 0.96 ± 0.33 points on the 9-point VAS and a prediction-outcome correlation coefficient of 0.74 (Fig. 1B).

The signature response increased nonlinearly with increasing stimulus intensity during thermal stimulation, but as expected, it was uniformly low for the pain-anticipation and pain-recall periods (Fig. 1C). To test the discrimination of painful from nonpainful warmth, we compared painful conditions ($>45^\circ\text{C}$, a temperature level that activates specific nociceptors¹⁷ and that was above the median temperature associated with reported pain) with warm conditions ($<45^\circ\text{C}$, which was below the median temperature associated with reported pain). Both sensitivity and specificity in the discrimination of pain from no pain were 94% or more for comparisons of pain versus nonpainful warmth, pain versus anticipation, and pain versus pain recall (Fig. 1D and Table 1).

Forced-choice tests showed 100% sensitivity and specificity for all three comparisons (Table 1), indicating that the signature response was always higher for painful stimulation than for anticipation or recall within an individual participant. In addition, the signature discriminated between relative differences in pain, with sensitivity and specificity of 93% or more when pain ratings differed by 2 or more points on the 9-point VAS (see the Supplementary Appendix). Thus, the neurologic signature was sensitive and specific to pain, with improved performance in the forced-choice test.

PAINFUL VERSUS NONPAINFUL HEAT

In study 2, the signature response increased monotonically across the six temperatures (Fig. 2A), with an expected nonlinear increase with temperature, and it correlated with both the reported level of pain ($r=0.73$) and the stimulus temperature ($r=0.65$). Signature responses increased with subjective intensity on a continuum across painful and nonpainful events (Fig. 2B), a finding that is consistent with contributions by colocalized wide-dynamic-range neurons and nocicep-

tive-specific neurons.¹⁷⁻¹⁹ However, mixed-effects regression analyses showed that the signature response increased more strongly with ratings of pain intensity than with ratings of warmth intensity ($\beta=0.66$, $t=2.58$, $P=0.02$) (Fig. 2B).

In trials involving painful heat, the neurologic signature strongly predicted pain intensity ($\beta=0.20$, $t=6.84$, $P<0.001$), even when we controlled for linear and nonlinear effects of temperature ($\beta=0.13$, $t=4.51$, $P<0.001$). In trials involving nonpainful heat, the neurologic signature weakly predicted warmth intensity ($\beta=0.06$, $t=2.04$, $P=0.08$) and did not predict warmth intensity after adjustment for temperature ($\beta=0.05$, $t=1.30$, $P=0.22$). These results suggest that the signature is related principally to the subjective sensation of pain but also reflects the overall intensity of somatic stimulation to some degree.

To assess discrimination performance, we averaged the neurologic signature response for painful conditions (rating, ≥ 100 ; mean rating, 138 points) and nonpainful conditions (rating, <100 ; mean rating, 60 points) for each participant. Because the field strengths of the scanners used in studies 1 and 2 differed (1.5 T vs. 3.0 T), we reestimated the signature-response threshold for painful versus nonpainful events, which was estimated to be 1.32 in study 2, as compared with 1.40 in study 1. The average signature response across trials accurately discriminated painful from nonpainful conditions with 93% sensitivity and specificity in the test of pain versus no pain (95% confidence interval [CI], 84 to 100 for both comparisons), and with 100% sensitivity and specificity (95% CI, 100 to 100) in the forced-choice test (Table 1).

The signature response also discriminated between clearly painful conditions and conditions near the pain threshold (mean score, 150 vs. 98 points) with 88% sensitivity (95% CI, 77 to 97) and 85% specificity (95% CI, 72 to 95) in the test of pain versus no pain and with 100% sensitivity and specificity in the forced-choice test. However, the signature response also discriminated between intense nonpainful warmth and mild nonpainful warmth (Table 1), suggesting that hyperalgesia or allodynia would be indicated by positive results of both the test of pain versus no pain and the forced-choice test.

Finally, tests of forced-choice discrimination across painful temperatures showed good performance, and tests across nonpainful temperatures showed poor performance, supporting the use of

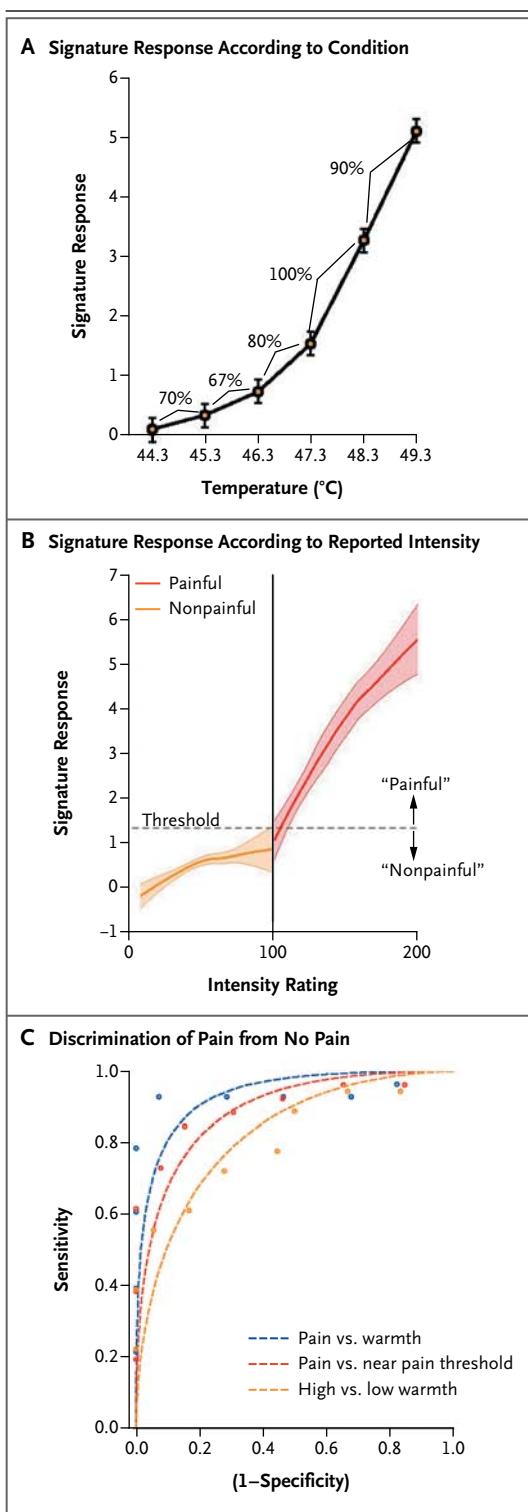


Figure 2. Application of the Neurologic Signature in Study 2.

Panel A shows the signature response across the temperatures used in study 2. The signature response was defined as the dot product of the signature-pattern weights from study 1 and the activation maps for each temperature within each individual participant. I bars show the standard error for the within-participant data. The signature response increased with increasing temperature, as did the level of reported pain. Percentages indicate the sensitivity and specificity for adjacent temperatures in the forced-choice classification. Sensitivity and specificity are equivalent for the forced-choice test and reflect the proportion of participants for whom the prediction based on the signature response was correct. Panel B shows the signature response as a function of reported intensity, for conditions rated as warm (nonpainful; orange) and those rated as painful (red). Loess smoothing was used to visualize the relationship; shaded areas show bootstrapped standard errors. The vertical line (at 100) divides conditions explicitly rated as painful from those rated as nonpainful, and the dashed horizontal line (at 1.32) is the classification threshold that maximizes the classification accuracy for painful versus nonpainful conditions. Panel C shows the discrimination performance for comparisons of pain and no pain. Performance (circles) was generally better than predicted by the gaussian model (dashed lines), suggesting a super-gaussian distribution of the signature response. Discrimination in the forced-choice test showed 100% sensitivity and specificity in all comparisons (data not shown).

for a temperature of 48.3°C versus 47.3°C, with 15 trials performed for each condition. However, performance dropped to near-chance levels when low temperatures were used (Fig. 2A, and the Supplementary Appendix).

SPECIFICITY OF NEUROLOGIC SIGNATURE FOR PHYSICAL PAIN

In study 3, comparisons of rejecter versus friend and pain versus warmth yielded similar levels of self-reported negative affect, and overlapping portions of many regions related to pain intensity were activated, including the bilateral anterior insula, medial thalamus, secondary somatosensory cortex, and dorsal posterior insula.⁶ These findings provided a good basis for a test of specificity.

The neurologic signature response was substantially stronger for physical pain than for any of the other conditions (warmth, rejecter, or friend) (Fig. 3A) and predicted pain ratings ($r=0.68$, $P<0.001$, with a mean prediction error of 0.84 points). As in study 1, the signature response predicted intensity ratings for noxious stimuli ($r=0.44$,

the signature to assess nociceptive responses. Sensitivity and specificity were 90% (95% CI, 81 to 97) for a temperature of 49.3°C versus 48.3°C, with only 4 trials performed at 49.3°C, and 100%

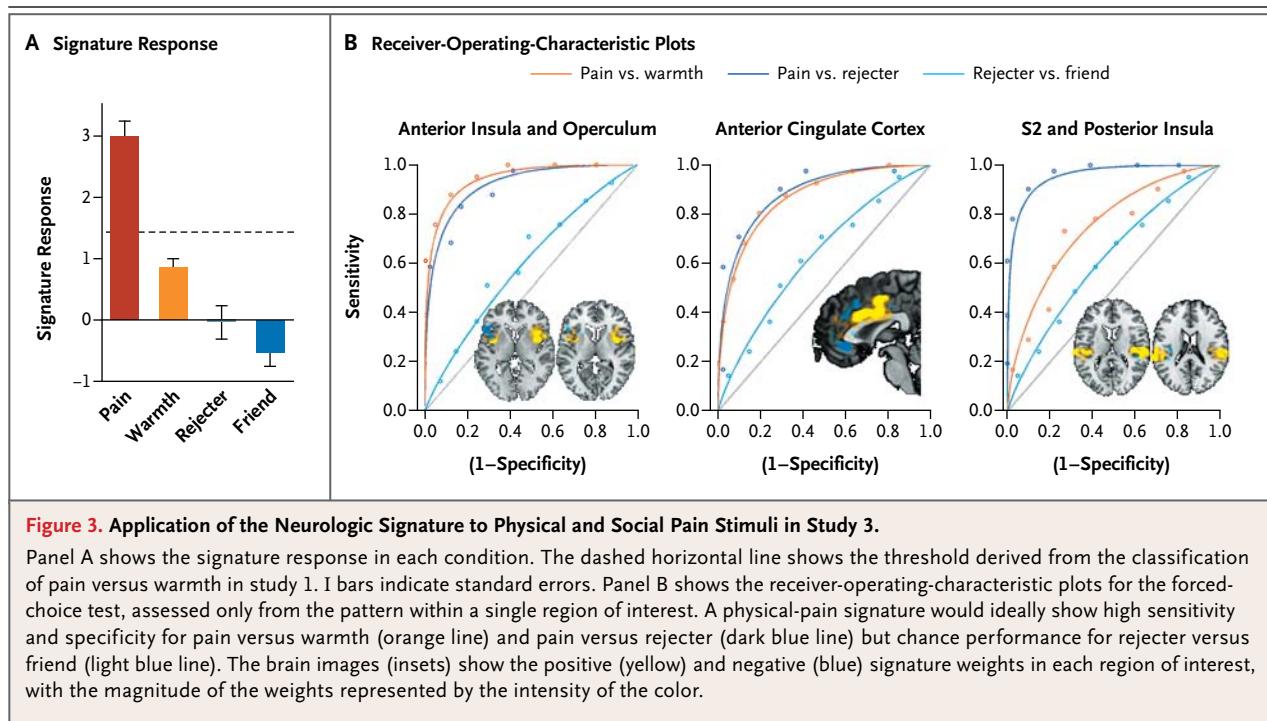


Figure 3. Application of the Neurologic Signature to Physical and Social Pain Stimuli in Study 3.

Panel A shows the signature response in each condition. The dashed horizontal line shows the threshold derived from the classification of pain versus warmth in study 1. I bars indicate standard errors. Panel B shows the receiver-operating-characteristic plots for the forced-choice test, assessed only from the pattern within a single region of interest. A physical-pain signature would ideally show high sensitivity and specificity for pain versus warmth (orange line) and pain versus rejecter (dark blue line) but chance performance for rejecter versus friend (light blue line). The brain images (insets) show the positive (yellow) and negative (blue) signature weights in each region of interest, with the magnitude of the weights represented by the intensity of the color.

$P < 0.01$) but not innocuous stimuli ($r = 0.02$, $P > 0.90$). With the use of the threshold derived from study 1, the response for the discrimination between pain and no pain had 85% sensitivity (95% CI, 76 to 94) and 78% specificity (95% CI, 67 to 89) for pain versus warmth and 93% sensitivity and specificity (95% CI, 86 to 98) for forced-choice discrimination, with similar performance for the comparison of pain and rejecter conditions ($P < 0.001$ for all comparisons) (Table 1). Discrimination between the rejecter and friend conditions was no better than would be expected by chance (Table 1).

This observed specificity may be driven by fine-grained differences in activity patterns in regions activated by both physical and social pain, an explanation that is consistent with the notion that different groups of neurons code for different affective events, or by differential activation of sensory-system-specific regions (e.g., the secondary somatosensory cortex for heat vs. the occipital cortex for images). If the first explanation holds, the pattern of activation, rather than the overall level of activation of a region, is the critical agent of discrimination.

To test these alternatives, we assessed the neurologic signature response derived from patterns within the dorsal anterior cingulate cortex, ante-

rior insula and operculum, and secondary somatosensory cortex and dorsal posterior insula individually (Fig. 3B, 3C, and 3D). Each region was activated by social pain (rejecter vs. friend) overall. However, in each region, the signature response reliably discriminated pain from the warm condition and pain from the rejecter condition (mean sensitivity and specificity in the forced-choice test, 78%) (Table S2 in the Supplementary Appendix) and performed at chance levels for the rejecter versus friend condition (mean sensitivity and specificity, 58%), suggesting that the pattern within these regions is critical for predicting pain.

REMIFENTANIL TREATMENT RESPONSE

Before drug infusion, in study 4, the signature response was greater for painful stimuli than for warm stimuli in both the open-infusion trials and the hidden-infusion trials ($t = 5.21$ and $t = 4.84$, respectively; $P < 0.001$ for both comparisons) (Fig. S6 in the Supplementary Appendix). During infusion, the signature response was reduced in parallel with increases in the drug effect-site concentration ($t = -2.78$ for trials with open infusion, and $t = -2.77$ for trials with hidden infusion; $P = 0.01$ for both comparisons).

At the maximum drug concentration, remifen-

tanil was associated with a reduction of 53% in the signature response, with no differences across the open and hidden infusions ($P=0.94$). The sensitivity and specificity for the discrimination between painful and warm stimuli in the forced-choice test were both 90% (95% CI, 79 to 100), with 95% sensitivity (95% CI, 86 to 100) and 62% specificity (95% CI, 43 to 79) in the test of pain versus no pain ($P<0.001$) (Table 1). Lower accuracy was expected because preinfusion signature responses in each condition were estimated from only three trials.

DISCUSSION

We identified an fMRI-based neurologic signature associated with thermal pain, discriminates physical pain from several other salient, aversive events, and is sensitive to the analgesic effects of opioids. This signature consisted of interpretable, stable patterns across regions known to show increased activity in association with experimentally induced pain, hyperalgesic or allodynic states,^{20,21} experimentally induced acute pain in patients,²² and experimentally induced tonic pain (pain caused by a stimulus of extended duration) in healthy persons.²³

The signature is distinguished from a general salience signal by its inclusion of somatic-specific regions, such as the ventrolateral thalamus,²⁴ the secondary somatosensory cortex,⁶ and the dorsal posterior insula,⁶ and by the identification of patterns of activity that are specific to physical pain within regions that are activated across many psychological processes (e.g., the anterior insula and the anterior cingulate cortex¹¹). Specificity to pain at the pattern level is consistent with findings that the anterior cingulate cortex and other association regions contain nociceptive-specific neurons as well as neurons with other properties²⁵ and that machine learning can identify fMRI patterns with specific functional properties.²⁶

The neurologic signature is predominantly bilateral but shows evidence of contralateral specificity in the primary and secondary somatosensory cortices, an observation that is consistent with previous work.^{15,27} These results build on previous studies^{16,28-31} by showing a signature that has more than 90% sensitivity and specificity for pain at the level of the individual person and that is consistently accurate across studies and scan-

ners. The forced-choice classification test we used could be translated into a test of hyperalgesia or allodynia in clinical studies, although the signature has not been validated for clinical pain and cannot currently be used in clinical tests.

If our findings are extended to clinical populations, brain-based signatures could be useful in confirming pain in situations in which patients are unable to communicate pain effectively or when self-reports are otherwise suspect. Such signatures could also help identify functional neuropathologic disorders³² that may underlie or confer a predisposition to chronic pain, even in the absence of overt structural lesions.³³ More broadly, brain-based signatures could accelerate the identification of neurophysiological subtypes of pain and intermediate markers for treatment discovery.³⁴ Such signatures could not, however, rule out the presence of pain with a nonnormative neurophysiological basis.

Before fMRI-based signatures for pain can be tested in medical decision-making settings, the generalizability of our findings must be assessed. Pain classification may be less accurate in patients than in healthy persons. Clinical use would require calibration across persons, scanning protocols, and research sites. Of the tests studied, the test of pain versus no pain is likely to be clinically useful in the broadest range of situations, but it is less strongly predictive than the forced-choice test. Finally, pain-associated fMRI patterns may differ according to body site,^{35,36} type of pain (visceral vs. cutaneous),³⁷ and clinical cause, potentially necessitating the development of multiple pain signatures. Nonetheless, our findings represent a step toward developing neurologic signatures for multiple types of pain and other cognitive and affective processes.

Supported by grants from the National Institute on Drug Abuse (1RC1DA028608 and R01DA027794), the National Institute of Mental Health (R01MH076136), and the National Science Foundation (0631637, to Dr. Wager).

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Drs. Ed Smith, John Jonides, Basil Margolis, Doug Noll, Robert Welsh, and Daphna Shohamy for helpful discussion and comments on an earlier version of the manuscript; Dr. Robert Whittington for participation in the drug administration and data collection in study 4; Damon Abraham and Kate Dahl for help with data collection; Dr. Tal Yarkoni for help with meta-analytic maps; Dr. Jason Buhle for assistance in planning study 2; Dr. Steven Shafer for assistance with pharmacokinetic modeling; and Dr. Lew Harvey for consultation on binary classification analyses.

REFERENCES

1. Institute of Medicine. Relieving pain in America: a blueprint for transforming prevention, care, education, and research. Washington, DC: National Academies Press, 2011.
2. Flor H, Turk D. Chronic pain: an integrated biobehavioral approach. Seattle: IASP Press, 2011.
3. Wager TD, Fields H. Placebo analgesia. In: McMahon SB, Koltzenburg M, eds. Wall and Melzack's textbook of pain. Oxford, United Kingdom: Churchill Livingstone (in press).
4. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer, 2009.
5. Mitchell T. Does machine learning really work? *AI Magazine* 1997;18:11.
6. Kross E, Berman MG, Mischel W, Smith EE, Wager TD. Social rejection shares somatosensory representations with physical pain. *Proc Natl Acad Sci U S A* 2011;108:6270-5.
7. Atlas LY, Whittington RA, Lindquist MA, Wielgosz J, Sonty N, Wager TD. Dissociable influences of opiates and expectations on pain. *J Neurosci* 2012;32: 8053-64.
8. Macdonald G, Leary MR. Why does social exclusion hurt? The relationship between social and physical pain. *Psychol Bull* 2005;131:202-23.
9. Minto CF, Schnider TW, Egan TD, et al. Influence of age and gender on the pharmacokinetics and pharmacodynamics of remifentanil. I. Model development. *Anesthesiology* 1997;86:10-23.
10. Wager TD, Atlas LY, Leotti LA, Rilling JK. Predicting individual differences in placebo analgesia: contributions of brain activity during anticipation and pain experience. *J Neurosci* 2011;31:439-52.
11. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods* 2011;8: 665-70.
12. Baliki MN, Geha PY, Apkarian AV. Parsing pain perception between nociceptive representation and magnitude estimation. *J Neurophysiol* 2009;101:875-87.
13. Lindquist MA, Meng Loh J, Atlas L, Wager TD. Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. *Neuroimage* 2009; 45 Suppl:S187-S198.
14. Wager TD, Rilling JK, Smith EE, et al. Placebo-induced changes in fMRI in the anticipation and experience of pain. *Science* 2004;303:1162-7.
15. Coghill RC, Sang CN, Maisog JM, Iadarola MJ. Pain intensity processing within the human brain: a bilateral, distributed mechanism. *J Neurophysiol* 1999;82: 1934-43.
16. Brodersen KH, Wiech K, Lomakina EI, et al. Decoding the perception of pain from fMRI using multivariate pattern analysis. *Neuroimage* 2012;63:1162-70.
17. Craig AD, Krout K, Andrew D. Quantitative response characteristics of thermoreceptive and nociceptive lamina I spinothalamic neurons in the cat. *J Neurophysiol* 2001;86:1459-80.
18. Dong WK, Salonen LD, Kawakami Y, Shiwaku T, Kaukoranta EM, Martin RF. Nociceptive responses of trigeminal neurons in SII-7b cortex of awake monkeys. *Brain Res* 1989;484:314-24.
19. Kenshalo DR, Iwata K, Sholas M, Thomas DA. Response properties and organization of nociceptive neurons in area 1 of monkey primary somatosensory cortex. *J Neurophysiol* 2000;84:719-29.
20. Meier ML, Brügger M, Ettlin DA, et al. Brain activation induced by dentine hypersensitivity pain — an fMRI study. *J Clin Periodontol* 2012;39:441-7.
21. Maihöfner C, Handwerker HO, Birklein F. Functional imaging of allodynia in complex regional pain syndrome. *Neurology* 2006;66:711-7.
22. Derbyshire SW, Jones AK, Creed F, et al. Cerebral responses to noxious thermal stimulation in chronic low back pain patients and normal controls. *Neuroimage* 2002;16:158-68.
23. Schreckenberger M, Siessmeier T, Viertmann A, et al. The unpleasantness of tonic pain is encoded by the insular cortex. *Neurology* 2005;64:1175-83.
24. DaSilva AF, Becerra L, Makris N, et al. Somatotopic activation in the human trigeminal pain pathway. *J Neurosci* 2002; 22:8183-92.
25. Sikes RW, Vogt BA. Nociceptive neurons in area 24 of rabbit cingulate cortex. *J Neurophysiol* 1992;68:1720-32.
26. Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 2005;8:679-85.
27. Bingel U, Quante M, Knab R, Bromm B, Weiller C, Büchel C. Single trial fMRI reveals significant contralateral bias in responses to laser pain within thalamus and somatosensory cortices. *Neuroimage* 2003;18:740-8.
28. Schulz E, Zherdin A, Tiemann L, Plant C, Ploner M. Decoding an individual's sensitivity to pain from the multivariate analysis of EEG data. *Cereb Cortex* 2012; 22:1118-23.
29. Brown JE, Chatterjee N, Younger J, Mackey S. Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation. *PLoS One* 2011;6(9):e24124.
30. Marquand A, Howard M, Brammer M, Chu C, Coen S, Mourão-Miranda J. Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *Neuroimage* 2010;49: 2178-89.
31. Rish I, Cecchi G, Baliki MN, Apkarian V. Sparse regression models of pain perception. In: Proceedings of the International Conference on Brain Informatics, Toronto, August 28-30, 2010.
32. Baliki MN, Petre B, Torbey S, et al. Corticostratal functional connectivity predicts transition to chronic back pain. *Nat Neurosci* 2012;15:1117-9.
33. Cohen SP, White RL, Kurihara C, et al. Epidural steroids, etanercept, or saline in subacute sciatica: a multicenter, randomized trial. *Ann Intern Med* 2012;156: 551-9.
34. Borsook D, Becerra L, Hargreaves R. A role for fMRI in optimizing CNS drug development. *Nat Rev Drug Discov* 2006; 5:411-24.
35. Brooks JCW, Zambrano L, Godinez A, Craig ADB, Tracey I. Somatotopic organisation of the human insula to painful heat studied with high resolution functional imaging. *Neuroimage* 2005;27:201-9.
36. Hua LH, Strigo IA, Baxter LC, Johnson SC, Craig ADB. Anteroposterior somatotopy of innocuous cooling activation focus in human dorsal posterior insular cortex. *Am J Physiol Regul Integr Comp Physiol* 2005;289:R319-R325.
37. Strigo IA, Duncan GH, Boivin M, Bushnell MC. Differentiation of visceral and cutaneous pain in the human brain. *J Neurophysiol* 2003;89:3294-303.

Copyright © 2013 Massachusetts Medical Society.

RECEIVE IMMEDIATE NOTIFICATION WHEN AN ARTICLE
IS PUBLISHED ONLINE FIRSTTo be notified by e-mail when *Journal* articles
are published Online First, sign up at NEJM.org.

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Wager TD, Atlas LY, Lindquist MA, Roy M, Woo C-W, Kross E. An fMRI-based neurologic signature of physical pain. *N Engl J Med* 2013;368:1388-97. DOI: 10.1056/NEJMoa1204471

Supporting Online Material for

An fMRI-based neurologic signature for physical pain

Tor D. Wager¹

Lauren Y. Atlas²

Martin A. Lindquist³

Mathieu Roy¹

Choong-Wan Woo¹

Ethan Kross⁴

¹ Department of Psychology and Neuroscience, University of Colorado, Boulder

² Department of Psychology, New York University

³ Department of Biostatistics, Johns Hopkins University

⁴ Department of Psychology, University of Michigan

Correspondence to:

Tor D. Wager

Department of Psychology and Neuroscience

University of Colorado, Boulder

345 UCB

Boulder, CO 80305

Email: tor.wager@colorado.edu

Telephone: (303) 492-7487

Table of Contents

Materials and Methods	p. 3
General Methods	p. 3
Study 1	p. 8
Study 2	p. 13
Study 3	p. 15
Study 4	p. 18
Supplementary Results	p. 20
Head movement analyses	p. 21
Time course of signature response	p. 22
Supplementary Figures	p. 24
Figure S1: <i>Head movement in Study 1</i>	p. 24
Figure S2: <i>Preprocessing and analysis stages</i>	p. 25
Figure S3: <i>Signature development in Study 1</i>	p. 26
Figure S4: <i>Correlation of signature response with the time course of objective stimulus delivery vs. reported pain in Study 1</i>	p. 27
Figure S5: <i>Modeling of drug effects in Study 4</i>	p. 28
Figure S6: <i>Signature response to remifentanil in Study 4</i>	p. 29
Supplementary Tables	p. 30
Table S1: <i>Peak coordinates from the machine learning analysis in Study 1</i>	p. 30
Table S2: <i>Forced-choice classification performance across studies</i>	p. 31
Table S3 <i>Pain/no-pain classification for selected single regions in Study 3</i>	p. 33
References	p. 35

Materials and Methods

Overview

A pattern of fMRI activity across brain regions—a neurologic signature—was optimized to predict physical pain. The signature’s sensitivity and specificity were tested in four separate studies:

Study 1 served as a “training set” for the signature. It involved the application of noxious thermal heat at temperatures calibrated to elicit ratings of non-painful warmth, low pain, medium pain, and high pain. Participants rated pain on every trial using a visual analogue scale. Sensitivity and specificity to physical pain were assessed using leave-one-subject-out cross-validated analyses.

Study 2 examined whether the signature trained on Study 1 predicts pain in the noxious range in new individuals from a new sample. In this study we again applied heat of varying temperatures, and participants made two ratings on each trial: 1) Whether the heat was warm or painful; and 2) How intense the stimulation was.

Study 3 provided a test of *signature specificity*. We examined physical pain processing and responses to social pain in participants who felt rejected after a recent a romantic breakup. Participants in this experiment received non-painful and high pain stimulation and made judgments about the pain on every trial. These participants also saw pictures of their ex-partners and close friends, which comprised the Social Pain contrast.

Study 4 examined the signature’s ability to discriminate painful from warm stimuli in a clinically relevant treatment context. It tested whether the signature responds to treatment with remifentanil, an opiate known to have analgesic effects. Participants received high and low thermal pain trials before, during, and after remifentanil was infused intraveneously.

In this Supplement, we present methodological details of data acquisition and analysis common across the four experiments (“General Methods”), followed by details unique to each individual experiment.

General Methods

Participants

All participants provided written informed consent. Studies were individually approved by the Columbia University Institutional Review Board. For all four studies, preliminary eligibility was assessed with a general health questionnaire, a pain safety screening form, and an fMRI safety screening form. Participants reported no history of psychiatric, neurological, or pain disorders. Ethnicity was assessed using self-report screening instruments prior to study procedures.

Thermal Stimulation and Pain Rating

In all four studies, thermal stimulation was delivered to the volar surface of the left (non-dominant) inner forearm applied using a TSA-II Neurosensory Analyzer (Medoc Ltd., Chapel Hill, NC) with a 16 mm Peltier thermode end-plate. Each stimulus lasted 8-12 seconds, depending on the Study, and always included a period of time during which the stimulus ramped up from baseline temperature [32°C] to the target temperature, and another steady ramp to baseline. The ramping was intended to help prevent head movement, and analyses described below confirmed that head movement does not increase at pain onset or during pain, and does not increase with increasing temperature (Supplementary Figure S1).

Before testing in Studies 1, 3, and 4, we performed a pain calibration procedure using methods described in previous work^{1,2}. In brief, we tested different sites on the forearm during calibration and used an adaptive staircase procedure to identify sites on the forearm with similar nociceptive profiles and to derive the individual participant’s dose-response curve for the relationship between applied thermal stimulation and reported pain (slope, intercept, R²). In Study 2, all participants received the same temperatures.

General fMRI Processing

FMRI data for all three studies were subjected to a standard series of preprocessing and analysis steps, which are shown in Supplementary Figure S2. The stages consisted of Preprocessing, Analysis, and Prediction/Evaluation. Preprocessing included a sequence of commonly used procedures performed using SPM software (Wellcome Trust Centre for Neuroimaging, London, UK). SPM5 was used for Studies 1, 3, and 4. SPM8 was used for Study 2, but the algorithms for all the steps we used were identical in both versions. Preprocessing also included several quality control procedures not typically performed in SPM per se, which were designed to be simple to implement (code can be obtained from wagerlab.colorado.edu or from the authors). Analysis consisted of a standard General Linear Model (GLM) analysis of each individual participant's data, and was conducted to summarize activity maps for painful heat and other conditions. Prediction involved estimating the signature response by computing the cross-product of these individual subject activation maps with a machine-learning signature pattern derived from other individuals. Specifically, the signature was derived from cross-validated machine learning analyses in Study 1 (see Signature Development below). It was applied to out-of-training-sample individual activity maps in Study 1 and new individual activity maps in Studies 2 and 3 to generate signature response values for each condition within each individual, which reflect a quantitative match to the pain signature pattern. Finally, evaluation involved quantifying the sensitivity and specificity of signature response to physical pain, and assessing the magnitude and significance of the opiate effect in Study 4.

These steps were employed for all analyses for all studies, except as noted below. Specifically, the initial Signature Development analyses involved several minor differences intended to ensure minimal artifacts in the data and minimize assumptions about the shape of the hemodynamic response to pain.

Preprocessing

Structural T1-weighted images were subjected to the following steps (Figure S2): Coregistration (SPM). We used SPM's iterative mutual information-based algorithm to coregister volumes to the mean functional image for each subject. Coregistration was manually checked by a trained analyst, and the starting point was adjusted and the algorithm re-run until the coregistration was satisfactory.

Warping to normative atlas (SPM). Structural images were normalized to MNI space using the generative Segmentation/Warping algorithm³ using the default parameters (7 x 8 x 7 nonlinear basis functions) and resliced to standard 2 x 2 x 2 mm voxels. Data were resampled to 3 x 3 x 3 mm voxels before signature development analyses (to facilitate efficient storage and processing) and before calculating signature response in all studies.

Functional images were subjected to the following steps (Supplementary Figure S2):

Outlier / gradient artifact detection (custom code). The purpose of this was to remove intermittent gradient and severe motion-related artifacts that are present to some degree in all fMRI data. On each individual scanning run, we identified image-wise outliers by computing both the mean and the standard deviation (across voxels) of values for each image for all slices. Mahalanobis distances for the matrix of slice-wise mean and standard deviation values

(concatenated) x functional volumes (time) were computed, and any values with a significant χ^2 value (corrected for multiple comparisons based on the more stringent of either false discovery rate or Bonferroni methods) were considered outliers (less than 1% of images were outliers). For each voxel, outlier time points were imputed with the voxel's overall run mean. Next, data across the entire run were Winsorized to three standard deviations. This procedure is similar to those commonly employed by many groups (e.g., http://www.nitrc.org/projects/art_repair/).

Slice-acquisition-timing correction (SPM). This interpolates the data to correct for differences in the acquisition time for each slice.

Image realignment (SPM). This is a rigid-body (6-parameter) registration to the mean functional image, and helps correct for head movement during scanning.

Percent signal change conversion (custom code). Time series data for each voxel were converted to percent signal change based on a spatially smoothed baseline time series (16 mm FWHM).

Warping to normative atlas (SPM). Warping parameters estimated from coregistered, high-resolution structural images were applied, and functional images were interpolated to 2 x 2 x 2 mm voxels.

Analysis

Except for machine learning analyses (see Signature Development below), activity maps for each condition within each participant were estimated using the GLM. For each individual, a set of

regressors was constructed for conditions of interest (e.g., heat at a particular temperature, aversive image presentation, etc.) using a stimulation epoch that lasted the duration of the event convolved with the canonical hemodynamic response implemented in SPM. The parameter estimates (regression slopes) for each condition thus provided an estimate at each voxel of the activation intensity for that condition. We also included a set of nuisance covariates designed to capture noise. These included, for each run: a) a constant term (intercept) for that run; b) dummy regressors for estimated outlier images from preprocessing, which varied in number depending on how many outliers were detected but was nearly always < 1% of images; and c) 24 movement-related covariates based on estimated movement during realignment, including 6 mean-centered motion parameter estimates, their squared values, their successive differences, and squared successive differences. Previous work has shown this to be helpful in reducing noise variance, violations of normality, and autocorrelation⁴.

Prediction and Evaluation

All assessments of performance were made at the level of the individual subjects, always based on a signature developed in other individuals using cross validation (Study 1) or simply applying the signature developed in Study 1 to new studies (Studies 2 and 3). For all tests, the signature response (BR) was estimated for each test subject in each test condition by taking the dot product of vectorized activation images ($\bar{\beta}_{map}$) with the signature pattern \bar{w}_{map} , i.e., ($BR = \bar{\beta}_{map}^T \bar{w}_{map}$), yielding a continuous scalar value. This value depends on the voxel size, but can be scaled based on the voxel volume (see Supplementary Discussion for additional scaling considerations). Values reported in this paper are for 27 mm³ voxels (i.e., 3 x 3 x 3 voxels). BR values derived from maps resliced to 2 x 2 x 2 mm voxels can be put on the same scale by multiplying by 27/8. We summarized the performance of the signature response in two ways: First, we assessed average prediction error (PE, the mean absolute deviation of predicted from observed pain ratings) when predicting continuous pain ratings. Second, we calculated sensitivity, specificity, positive predictive value, and effect sizes related to binary classification. We assessed binary classification decisions for painful stimulation relative to non-painful warmth, pain anticipation, pain recall, and social pain-inducing events.

We performed two kinds of binary classification tests. In the pain/no-pain test, sensitivity is the probability of a positive test—i.e., that the signature response was above a given criterion

threshold—given that a person experienced pain (vs. one of the comparison conditions below). Specificity is the probability of a negative test given that a person experienced a condition other than pain. Positive predictive value is the probability that pain (vs. a comparison condition) was experienced given a positive test result. Effect size provides a continuous measure of the ability of the signature to separate pain from a comparison condition, and is reported as both (1) d_a , a measure of the distance between the mean signature response in the pain-present vs. pain-absent conditions, divided by their pooled standard deviation, and (2) the area under the Receiver Operating Characteristic (ROC) curve (AUC), estimated directly using numerical integration of the ROC under all threshold values that yielded unique sensitivity/specificity values (0.5 is chance, and 1 is perfect discrimination). In the forced-choice discrimination test, signature response is compared for two conditions tested within the same individual, and the higher is chosen as more painful. In the forced-choice test, the ROC curves are symmetrical, and sensitivity, specificity, and positive predictive value are equivalent to each other and to decision accuracy (i.e., the probability with which the more painful of the two conditions is selected).

The forced-choice test has several advantages that make it particularly useful in the fMRI setting. First, the forced-choice test is ‘threshold free’ in the sense that an absolute decision threshold across individuals is not required; zero is used as the threshold for the difference between the two paired alternatives. Thus, individual differences in the shape and amplitude of the blood oxygen level dependent (BOLD) fMRI response^{5,6} do not add noise in this kind of test. In addition, as the amplitude of the BOLD response varies as a function of field strength and scanner noise, the threshold in the pain/no-pain test must be calibrated for different scanners and field strengths (see, e.g., the thresholds for Study 1, collected at 1.5 T, vs. Study 2, collected at 3.0 T, in Table 1 in the main text). Second, the forced-choice test likely provides a more realistic assessment of the signature’s performance for validation purposes. Prediction error and sensitivity/specificity in the tests is calculated assuming that pain reports always accurately reflect experienced pain intensity in the normative samples we test here (i.e., a person reporting a “5” on the visual analogue scale always experiences more pain than a person reporting a “4”). However, this may not always be the case. Individuals may use the rating scale in somewhat different ways (e.g., the same experience may be reported by one person as a “5” on the visual analogue scale and by another as a “4”), which can reduce the apparent performance of even a perfect diagnostic test. Forced-choice discrimination performance does not require this

assumption, as two conditions are compared within the same individual. The only condition that must hold for the ‘ground truth’ to be accurate is that an individual’s pain reports must increase monotonically with pain experience; more pain should be reported as more painful.

Study 1

Participants

Twenty healthy, right-handed participants completed the study ($M_{age} = 28.8$ years, 8 female). The sample consisted of 79% Caucasian, 5% Hispanic, and 16% African American participants. Data were collected between 2005-2006.

Materials and Procedures

fMRI task design

fMRI images were acquired during 8 functional runs (8 trials/run, 64 trials). The thermode was placed on a different skin site for each run, with two total runs per skin site, and 12 trials at each of 4 target pain intensities—non-painful warmth (Level 1), low pain (Level 3), medium pain (Level 5), and high pain (Level 7)—were delivered across the runs. Temperatures were selected for each individual based on a thermal pain calibration procedure (see above, “Thermal stimulation and pain ratings”). At the start of each trial, a square appeared in the center of the screen for 50 ms, followed by the presentation of a cue. The cue consisted of a male or female face showing a happy or fearful expression (33 ms) followed by a mask consisting of the same face presented for 1467 ms. Participants were not aware of the type of emotional face presented, and all analyses collapse across the different face types to examine brain activity as a function of temperature and reported pain. Effects of faces will be reported in a later manuscript.

During each trial, cues (2 sec) were followed by a six-second anticipatory interval during which a fixation cross was presented on the screen. Then, thermal stimulation was delivered at one of the four intensities, followed by a 14 sec rest interval during which participants fixated on a cross. The words “How painful?” then appeared on the screen for four seconds above a 9-point visual analogue scale (VAS), and participants rated the intensity of the stimulus using an fMRI-compatible track-ball (Resonance Technologies, Inc.) Continuous responses were recorded, with resolution equivalent to the screen resolution (approximately 600 discrete values).

fMRI Acquisition and Analysis

Image acquisition. Whole-brain fMRI data were acquired on a 1.5T GE Signa Twin Speed Excite HD scanner (GE Medical Systems) at Columbia University's Program for Imaging in Cognitive Science (PICS). Structural images were acquired using high-resolution T1 spoiled gradient recall images (SPGR) for anatomical localization and warping to a standard space. Functional images were acquired with an echo-planar imaging sequence (EPI; TR = 2000 ms, TE = 34 ms, field of view = 224 mm, 64x64 matrix, 3.5 x 3.5 x 4.0 mm voxels, 29 slices), and were resliced to 3 x 3 x 3 mm voxels after inter-subject normalization. Each run lasted 6 minutes and 18 seconds (189 TRs). Stimulus presentation and behavioral data acquisition were controlled using E-Prime software (PST Inc.).

Preprocessing. Preprocessing was identical to that described in the General Methods, except that a) an additional denoising step was used to minimize artifacts for signature development, and b) FSL software was used for realignment. Denoising used a component-based strategy similar to published work^{7,8}. We estimated the first 10 principal components (PCs) on the images from each scanning run, before any other processing. We constructed a task-related design matrix with the trial onsets convolved with the canonical HRF (no temperature information was entered to avoid bias), and a nuisance-related design matrix based on head movement parameters and outlier time points identified as described above. Components that appeared clearly artifactual (e.g., those expressed only at the edge of the brain, those that included an obvious single spike, etc.) and were related to the nuisance regressors but not the task, were removed (1.06 ± 0.59 (S.D.) components per run). Analyses of Studies 2 and 3 did not involve this step, and future studies are needed to assess the benefits of this manual procedure.

Signature development analysis. Signature development analyses were conducted on Study 1 using custom Matlab code (see 14) implementing LASSO-PCR, a cross-validated, regularized regression procedure. LASSO, or Least Absolute Shrinkage and Selection Operator-regularized regression⁹, was implemented in Matlab by Guilherme Rocha and Peng Zhao. This was embedded within a leave-one-subject out cross-validation loop that first used principal components-based data reduction so that selection was performed on components, as described in previous work¹⁰. The resulting pattern of regression weights constituted the signature, which was applied to average pain maps and general linear model-based activation maps in Studies 1-3. All predictions made for Study 1 data were cross-validated (see below).

The signature development analysis consisted of five steps: 1) Feature selection: Voxels within an a priori mask of pain-related brain regions was selected based on prior literature; 2) Data averaging: Data during pain from each in-mask voxel were averaged within each stimulus intensity for each individual, to generate 4 pain-related activation maps per individual; 3) Machine learning: LASSO-PCR was run using those maps to predict pain reports; 4) Bootstrapping was used provide P-values for voxel weights in order to threshold the signature weights for display and interpretation; and 5) Permutation tests were used to validate the unbiased nature of the procedure.

1. Feature selection. To accomplish Step (1), the automated meta-analysis toolbox Neurosynth (www.neurosynth.org) was used to create a mask based on a meta-analysis of previous studies that frequently use the word ‘pain’ to select voxels a priori¹¹. The mask (see Figure S3A, top) was based on regions showing consistent results across 224 published studies (out of 4,393 total studies in the database) in a ‘reverse inference’ analysis, which was a χ^2 analysis of the 2×2 contingency table of counts of [activated (within 10 mm) vs. non-activated] \times [pain vs. non-pain] within each voxel. Studies were counted as involving ‘pain’ if they mentioned ‘pain’ more than 1 time per 1000 words in the study (the default value in neurosynth) and thresholded at $q < 0.05$ False Discovery Rate ($P < 0.0072$) corrected. The mask included 22,379 positive voxels ($2 \times 2 \times 2$ mm, resliced to $3 \times 3 \times 3$ mm for analysis) in which activity positively predicted pain (6.35% of the volume of the standard SPM5/8 brain mask brainmask.nii) and 10,940 negative voxels in which activity negatively predicted pain (3.1%), for a total of 9.45% of the in-brain volume. Weights from all voxels in this mask were used to estimate signature response and make predictions (no further thresholding was used for predictive purposes).

2. Data averaging. To accomplish Step (2), we averaged data within each trial in each voxel over the period 8-24 seconds after heat onset, and then averaged across the 12 trials for each stimulus intensity. This time window was chosen a priori based on the approximate time when reported pain is high from previous work¹²⁻¹⁵; which is later than typical responses for a similar stimulation epoch due to temporal summation and hemodynamic lag in pain-related activity^{13,16}. Simple averaging has the advantage of simplicity and lack of strong assumptions about the shape of the hemodynamic response, although improvements in the use of timing information is a rich direction for future improvement that has already started to be explored¹⁷.

3. Machine learning. To accomplish Step (3), we used cross-validated LASSO-PCR with activation maps from each condition within participants as the predictor, and average pain reports from each condition within participants as the outcome. The linear algorithm provided interpretable brain maps^{10,18} composed of linear weights on voxels, which is a substantial advantage over nonlinear kernel methods. We did not explore nonlinear methods here.

We used leave-one-subject-out cross-validation to estimate prediction error (PE; mean absolute deviations between predicted and actual temperatures) on new trials. This standard approach in machine learning involves dividing the sample into a training set (all but one participant) and a test set (the test participant). LASSO-PCR was used to estimate regression weights for each voxel from the training dataset (\bar{w}_{map} , the signature pattern), and then predictions were made for the test participant by and taking the dot-product of the test brain activation maps ($\bar{\beta}_{map}$) and the signature pattern ($\bar{\beta}_{map} \bullet \bar{w}_{map}$). This yielded a scalar predicted pain value (the signature response) for each condition, and prediction error was quantified. The procedure was repeated 20 times (once for each participant) so that each trial was part of the test set exactly once. This procedure yields minimally biased estimates of prediction accuracy for new participants (there is a slight bias in accuracy towards zero, as with all cross-validation methods). Weight maps applied to Study 1 were always based on data from out-of-test-sample individuals, and the final signature weights (applied to Studies 2-4) were based on the full Study 1 sample.

To apply the signature to new activation maps across multiple conditions (i.e., anticipation, stimulation, and pain recall at each intensity, and other maps in Studies 2-4), we used a standard general linear model (GLM) with the canonical SPM hemodynamic response function to simultaneously estimate activation maps ($\bar{\beta}_{map}$) for each condition, and then applied the signature pattern ($\bar{\beta}_{map} \bullet \bar{w}_{map}$) to yield a scalar signature response value for each condition. The signature response values are thus predictions of the magnitude of pain for a given condition, and their values across conditions can be compared and tested.

In our initial analyses of Study 1, we compared LASSO-PCR results with those from another popular method, Support Vector Regression (SVR;¹⁹) in order to check whether predictions were similar and whether SVR produced similar accuracy levels. Predictions and accuracy levels were nearly identical with SVR in all cases (predictions between LASSO-PCR

and SVR were correlated $> r = 0.99$ in most cases), so we do not focus on the SVR results. We prefer the LASSO-PCR results for transparency and consistency with our previous work. LASSO-PCR and SVR produced very similar results in all analyses we performed, and we do not consider the choice of algorithm to be critical, though algorithms that yield improved results could be developed.

4. Bootstrap tests. To accomplish Step (4) and threshold voxel weights for interpretation and display, we constructed 5,000 bootstrap samples (with replacement) consisting of paired brain and outcome data and ran LASSO-PCR on each. Two-tailed, uncorrected P-values were calculated for each voxel based on the proportion of weights below or above zero, as in previous work (1, 20), and subjected to False Discovery Rate correction ($P < 0.0028$, 355 significant voxels; Supplementary Figure S3B, C). The signature weight map applied to Studies 1-3 for diagnostic purposes was not thresholded; all weights were used.

5. Permutation tests. To accomplish Step (5), we permuted the data 5,000 times, repeating the cross-validated LASSO-PCR analysis for each permuted dataset. The correlation between predicted and observed pain should be symmetrically distributed around zero if the procedure is unbiased, and this was tested and confirmed (Supplementary Figure S3D). In addition, the mean prediction error and predicted pain-observed pain correlation were far lower and higher, respectively, for the correct permutation ($P < 0.001$ for both; Figure S3D), demonstrating that the prediction results were far better than what would be expected by chance.

Study 2

Participants

Thirty-three healthy, right-handed participants completed the study ($M_{age} = 27.9 \pm 9.0$ years, 22 females). The sample consisted of 39% Caucasian, 33% Asian, 12% Hispanic, and 15% African American participants. Data were collected between 2010-2011.

Materials and Procedures

Thermal stimulation and pain ratings

Thermal stimulation was delivered to locations on the left volar forearm that alternated between runs. Each stimulus lasted 12.5 seconds, with 3-second ramp-up and 2-second ramp-down periods and 7.5 seconds at target temperature. Trials at six discrete temperatures were administered (level 1: 44.3°C, level 2: 45.3°C, level 3: 46.3°C, level 4: 47.3°C, level 5: 48.3°C,

level 6: 49.3°C). After each stimulus, participants rated explicitly whether it was painful or not. If they rated it as non-painful, they were then prompted to rate warmth intensity on a 100-point VAS anchored with “no sensation at all” and “very warm but not yet painful.” If they rated it as painful, they rated pain intensity on a 100-point VAS anchored with “no pain” and “worst imaginable pain.”

fMRI task design

fMRI images were acquired during 10 functional runs. Runs 1, 2, 4, 8 and 9 were “standard” runs, during which were delivered a total of 11 stimulations from each of levels 1-5, for a total of 55 stimuli. Transitional frequencies were counterbalanced so that each temperature was preceded twice by each of the five temperatures and each run started with a different temperature. Different presentation orders were generated for each participant. On Runs 5-6 temperatures were increased one degree, with 4 stimuli at each of levels 2-6. During two additional runs (not analyzed here), participants were instructed on the use of mental imagery to modify pain. These are beyond the scope of the current paper and will be presented elsewhere.

Each trial consisted of a stimulus (12.5 sec), a 4.5-8.5 sec delay, a 4 sec painful/non-painful decision period (participants pressed the left or right button on the side of an MR-compatible trackball), a 7-sec continuous warmth or pain rating period (VAS ratings were made using the trackball and confirmed with a button-press), and 23-27 sec of rest. During both rest and stimulation, participants fixated on a cross presented on-screen.

fMRI Acquisition and Analysis

Imaging acquisition. Whole-brain fMRI data were acquired on a 3T Philips Achieva TX scanner at the PICS Center. Structural images were acquired using high-resolution T1 spoiled gradient recall images (SPGR) for anatomical localization and warping to a standard space. Functional EPI images were acquired with TR = 2000 ms, TE = 20 ms, field of view = 224 mm, 64x64 matrix, 3 x 3 x 3 mm voxels, 42 interleaved slices, parallel imaging, SENSE factor 1.5. Runs lasted between 6:22 and 6:58 (191 or 209 TRs). Stimulus presentation and data acquisition were controlled using E-Prime.

Preprocessing and analysis. Image preprocessing and analysis were performed as described under General fMRI Processing above. First-level GLM analyses for each participant included regressors for stimulation periods for each of the 6 levels and the 11-sec rating periods,

linear drift across time within each run, and indicator vectors for outliers and head movement as described above. The signature pattern from Study 1 was used to estimate the signature response for each participant in each condition, and these values were used in binary classification analyses.

To assess classification performance for painful vs. non-painful trials, we averaged signature responses for non-painful and painful trials, and subjected these average responses to sensitivity/specificity analyses. Because this study was collected on a different scanner with a higher field strength, signature responses were on a different scale and a different classification threshold was determined for pain/no-pain classification. Forced-choice analyses are threshold-free and do not require this adjustment.

Regression models. In a second model, we included separate regressors for each individual trial, and applied the signature pattern from Study 1 to estimate the signature response for each individual trial. We used these values in mixed effects regression models predicting pain and temperature. Both warmth ratings and pain ratings were very sensitive to temperature increases: Pain ratings increased 20.8 ± 12.9 (SD) units/ $^{\circ}\text{C}$, and warmth ratings increased 17.7 ± 12.7 units/ $^{\circ}\text{C}$.

In the regression analyses, we tested models in which we assessed performance in predicting pain controlling for temperature. To completely control for temperature, we included covariates that controlled for all possible pairwise differences between temperatures (level 6 vs. 5, 5 vs. 4, 4 vs. 3, 3 vs. 2, and 2 vs. 1), thus controlling for temperature estimated in a nonparametric fashion, without assuming linearity. This analysis removed much of the variation in pain report (as most of the variance was caused by temperature), but served as a test of whether signature responses predicted pain even when completely accounting for the effects of heat itself.

Study 3

Participants

Forty right handed, native English speakers (21 females, $M_{age} = 20.78$, SD = 2.59) gave informed consent. All participants experienced an unwanted romantic relationship break-up within the past six months ($M = 2.74$ months; SD = 1.70 months), and indicated that thinking about their break-up experience led them to feel rejected. All participants scored above the

midpoint on a 1 (not at all rejected) to 7 (very rejected) scale that asked them to rate how rejected they feel when they think about their rejection experience ($M = 5.60$, $SD = 1.06$). The sample consisted of 60% Caucasian, 20% Asians, 10% African Americans, and 10% other ethnicities. Data were collected between 2007-2008. Data on the basic group activation maps for physical and social pain contrasts were published previously²⁰, but the analyses and substantive conclusions were different from and complementary to those reported here.

Materials and Procedures

Social Pain Stimuli

The social rejection task was modeled after (a) fMRI research that used photographs provided by participants to elicit powerful emotions, including maternal love, romantic love, and rejection (21-24) and (b) behavioral research indicating that cueing people to recall autobiographical rejection experiences is an effective way of reactivating social rejection related distress (e.g., 25, 26, 27). The stimuli for this task consisted of: (a) a headshot photograph of each participant's ex-partner and a same gendered friend with whom they shared a positive experience around the time of their break-up ($M = 2.46$ months; $SD = 1.70$ months), and (b) cue phrases appearing beneath each photograph which directed participants to focus on a specific experience they shared with each person.

All photographs were cropped so that the total area of the photograph taken up by the face was constant across ex-partner and friend images ($t = 1.42$, $P = .16$). To be sure that the photographs participants provided were matched in terms of picture quality, we had a group of ten individuals who were blind to the study goals and hypotheses rate the picture quality of each photograph. Ex-partner and friend photographs did not differ significantly on this dimension ($t = 1.32$, $P = .20$). Judges also rated the attractiveness level of the individuals depicted in ex-partner and friend photos, which also did not differ significantly ($t = .89$, $P = .38$).

When participants viewed the photograph of their ex-partner during the social rejection task they were instructed to think about how they felt during their specific break-up experience; when they viewed the photograph of their friend they were instructed to think about how they felt during their recent positive experience with that person. To help participants focus on these specific experiences during the task we included a short cue phrase beneath each photograph (e.g., "rejected by Marc"; "party with Ted"). Participants generated these cue phrases on their own, prior to the day of scanning using a procedure developed in prior research²¹. Specifically,

they first wrote about their specific break-up experience with their ex-partner and their specific positive experience with their friend. Subsequently, they were asked to create a cue phrase that captured the gist of their experience. They were reminded of the cues they generated and their break-up experiences on the day of scanning following established procedures²¹.

Physical Pain Stimuli

As in Study 1 and prior research^{2,14,22}, a calibration procedure was used to select heat intensities that participants judged to be non-painful (“warm,” Level 2 on a 10-point scale) vs. near the limit of pain tolerance (“hot,” as close as possible to Level 8 on a 10-point scale, though intensity was capped at 48°C). The mean low temperature for the sample was 39.9°C (SD = 2.76°C); the mean high temperature was 46.6°C (SD = 1.72°C). In the scanner, participants rated both physical and social pain on a 5-point scale using a five-button unit under their right hand, with lower numbers reflecting more distress.

Task Training

Prior to scanning, the experimenter walked participants through each step of the social rejection task (referred to as the “photograph” task to participants) and the physical pain task (referred to as the “heat” task to participants). They were told that during the “photograph” task they would see the photographs of their ex-partner and friend. The experimenter explained that beneath each photograph the cue-phrases they generated earlier would appear. When they saw each photograph they were asked to look directly at it and think about how they felt during the specific experience associated with the cue-phrase. Thus when participants viewed the photograph of their ex-partner they were directed to think about how they felt during their break-up experience with that person; when they viewed the photograph of their friend they were directed to think about how they felt during their positive experience with that person. During the physical pain task, participants were instructed to focus on the fixation cross that appeared on the screen during the trials, and think about the sensations they experienced as the thermode on their arm heated up. They were then instructed how to rate their affect after each type of trial, and how to perform the visuospatial control task.

fMRI Acquisition and Analysis

Acquisition

Whole-brain functional data were acquired on a GE 1.5 T scanner at the PICS Center (the

same scanner used in Study 1) in 24 contiguous axial slices (4.5 mm thick, 3.5 x 3.5 mm in-plane resolution) parallel to the anterior commissure-posterior commissure (AC- PC) line with a T2*-weighted spiral in out sequence (repetition time [TR] = 2000, echo time [TE] = 40, flip angle = 84, field of view [FOV] = 22.4) in 4 runs of 184 volumes each (368 sec each). Structural data were acquired with a T1-weighted spoiled gradient recalled echo scan (180 slices, 1 mm thick, in-plane resolution 1 x 1 mm; TR = 19, TE = 5, flip angle = 20, FOV = 25.6).

Analysis

Image preprocessing and analysis were performed as described under General fMRI Processing above, except that functional data were smoothed with a 6 mm FWHM Gaussian kernel after spatial warping and prior to analysis (as done in a prior publication on these data;²⁰). First-level GLM analyses for each participant included regressors for Rejector photos, Friend photos, Hot (painful) stimulation, and Warm (peri-pain threshold) stimulation periods, as well as covariates for the 5 sec affect rating periods for each condition and movement and outlier covariates for each run. The signature pattern from Study 1 was used to estimate the signature response for each participant in each condition, and these values were used in binary classification analyses.

Study 4

Participants

Twenty-one healthy, right-handed participants completed the study ($M_{age} = 24.7 \pm 4.18$ years, 11 females). The sample consisted of 40% Caucasian, 15 % Asian, 30% Hispanic, and 15% African American participants. Data were collected between 2007-2008. Data on dissociable drug effects and expectancy effects were published previously²³, but the analyses and substantive conclusions were different from and complementary to those reported here.

Materials and Procedures

Thermal stimulation and pain ratings

FMRI images were acquired during 2 functional runs of 6 blocks each (6 trials/block, 64 trials), with 30-second breaks between blocks, during which an experimenter rotated the thermode location. The thermode was placed on a different skin site for each block, and skin sites were stimulated in the same order on each run. Temperatures were selected for each individual based on a thermal pain calibration procedure (see above, “Thermal stimulation and

pain ratings”), and thermal stimulation alternated between stimuli calibrated to elicit low pain (Level 2; $M = 41.16^{\circ}\text{C}$, $SD = 2.64$) and high pain (level 8; $M = 47.05^{\circ}\text{C}$, $SD = 1.69$).

Remifentanil administration and experimental design

During fMRI scanning, participants received remifentanil hydrochloride (Ultiva; Mylan Institutional) intravenously under two conditions (‘runs’): Open administration, in which participants were fully informed about the drug infusion, and Hidden administration, during which participants were told they would receive no drug. Remifentanil administration proceeded identically in both runs. Run order was counterbalanced, such that half the participants received the Open run first, and half the Hidden run first, in a crossover design. Participants received remifentanil at doses individually selected to elicit pain relief without sedation, based on a pre-experiment dosing procedure. The average dose administered was $0.043 \mu\text{g}/\text{kg}/\text{min}$ ($SD = 0.01$). Remifentanil infusion began after the first block (before trial 7), and infusion proceeded steadily throughout blocks 2-4, for the next 18 trials. Infusion was stopped and a washout period began following the fourth block, and anatomical images were acquired between runs to allow additional time, so that the brain concentrations of remifentanil were negligible at the start of the next run.

Thirty-six trials were administered in each run, 18 with painful heat and 18 with non-painful warmth. Pain and warm trials alternated, with order (pain first or warm first) counterbalanced across participants in a crossover design. At the start of each trial, participants heard an auditory tone (an orienting cue) and saw the words “warm” or “hot” on the screen for 3 s. Following a 7-13s jittered anticipation interval ($M = 10.16$ s, $SD = 2.64$), participants felt heat from the thermode at temperatures calibrated to elicit either low or high pain (1.5s ramp-up, 7s at peak, 1.5s ramp-down). This was followed by a 9-15 s rest interval ($M = 11.67$ s, $SD = 2.50$), during which participants fixated on a cross. The words “How painful?” then appeared on the screen for 4 - 6 seconds above a 9-point visual analogue scale (VAS), accompanied by an orienting tone. As in Study 1, participants rated the intensity of the stimulus using an fMRI-compatible track-ball (Resonance Technologies, Inc.). The next trial began after 9-15s ($M = 11.46$ s, $SD = 2.57$).

fMRI Acquisition and Analysis

Image acquisition. Whole-brain structural (T1-weighted SPGR) and EPI fMRI data were acquired on a 1.5T GE Signa Twin Speed Excite HD scanner (GE Medical Systems) at Columbia

University's Program for Imaging in Cognitive Science (PICS), as in Studies 1 and 3. (EPI; TR = 2000 ms, TE = 34 ms, field of view = 224 mm, 64x64 matrix, 3.5 x 3.5 x 4.0 mm voxels, 28 slices). Each run lasted 33 minutes and 20 seconds (1000 TRs), divided into six blocks, with a brief pause between blocks 4 and 5 to prevent scanner overheating. Stimulus presentation and behavioral data acquisition were controlled using E-Prime software (PST Inc.).

Preprocessing. Preprocessing was identical to that described in the General Methods, except that FSL software was used for realignment.

Analysis. We used first-level (single-subject) GLM regression parameter estimates from our previously published study²³ (but adjusted to 3 x 3 x 3 mm voxels), which maintained consistency in modeling of the events and drug effects across the previous report and this one. Full details of the model are provided in the previous publication²³, but in brief, we modeled effects of painful (Hot) and non-painful (Warm) stimulation in each of Open and Hidden runs with separate regressors. To model drug effects across time, we used a pharmacokinetic model and parameter estimates based on age, weight, and sex^{24,25} to estimate the drug effect site concentration second-by-second during drug infusion. Those values were normalized to a peak amplitude of 1 and used to create a “parametric modulator” regressor for each condition, which is orthogonal to the average regressor across trials and estimates changes in heat-evoked responses across time that are linearly related to drug effect site concentration. Example regressors for Hot trials are shown in Supplementary Figure S5A. The regressor capturing the average response across trials is shown in green, and the drug concentration regressor is shown in red. To capture additional effects of expectations and other time-varying effects that do not follow the time-course of drug effects, we included an additional parametric modulator, which modeled the period of infusion vs. pre- and post-infusion baseline, orthogonalized to the drug effect site concentration regressor. This is shown in blue in Supplementary Figure S5. Together, the regressors capture a range of modulatory effects across time, including drug effects based on the pharmacokinetic model (Supplementary Figure S5B).

To test Hot vs. Warm and drug effects on the signature response, we applied the signature pattern from Study 1 to each regression parameter estimate ($\bar{\beta}_{map}$) map to yield a single amplitude value (BR) for each regressor within each participant. The significance of the drug modulation effect on signature response was tested by conducting a t-test on the BR values for the drug effect site concentration regressor. To visualize the responses (Figure 4, Supplementary

Figure S5C), we reconstructed the fitted responses for Hot and Warm trials in each of Open and Hidden administration by multiplying the appropriate regressors in the design matrix X by BR for each participant. This yielded an overall fitted time course for each condition within each subject (averages across participants are shown in Supplementary Figure S5C). To conduct analyses on pre-drug infusion and peak drug infusion trials, we constructed a GLM design matrix with regressors for each trial, and used it to estimate the amplitude of the fitted response on each trial. Those estimates, averaged across participants, are shown by the solid line in Supplementary Figure S5C. Estimates for pre-drug infusion trials were obtained by averaging across amplitudes for Trials 1-3 for each participant, and estimates for peak drug infusion trials were obtained by averaging amplitudes for Trials 10-12.

Supplementary Results

The following analyses examine several methodological aspects of the study, and presented as supporting information. They demonstrate that a) head movement is not induced by thermal stimulation and does not drive pain-predictive results; and b) the time course of the signature response tracks pain experience more closely than the time course of noxious heat itself.

Head movement analyses

In Study 1, to assess whether noxious thermal stimulation caused head movement, we quantified relationships between head movement and time within trial (anticipation, stimulation, and rating periods). We estimated head movement by taking the absolute successive differences between motion estimates from rigid-body image realignment during preprocessing. For each of the six directions of potential movement (lateral, anterior-posterior, and inferior-superior translation and roll, pitch, and yaw), movement was highest at the onset of the pain-predictive cue, but was still within standard tolerances even for the worst movement direction (< 0.08 mm / 0.06 degrees; Supplementary Figure S1A/B). Movement dropped within a few seconds to low levels, and stayed low throughout the stimulation epoch without responding to heat onset or offset. We also averaged head movement during the stimulation epoch as a function of stimulus temperature. Mixed-effects regression analyses revealed no significant relationships between temperature and head movement for any parameter (Supplementary Figure S1C/D). Effect sizes

ranged from $Z = 0.17 - 0.92$, all $P > 0.10$. Similar results were obtained for other studies, but are omitted here for brevity.

We also quantified the degree to which head movement and the inclusion of movement-related covariates impacted the sensitivity/specificity analyses. If pain is correlated with head movement, including head movement-related covariates should reduce performance in discriminating pain from other conditions. Conversely, if it is unrelated, controlling for head movement may increase discrimination accuracy by removing noise in the fMRI data. Across the six analyses of sensitivity/specificity reported for Study 1 (Pain vs. Low pain, Pain vs. Anticipation, and Pain vs. Pain Recall for each of pain/no-pain discrimination and forced-choice discrimination cases), effect sizes were moderately larger when controlling for head movement as described above (difference in $d_a = 0.03 - 0.83$, mean = 0.49). Similar results were obtained for other studies, but are omitted here for brevity.

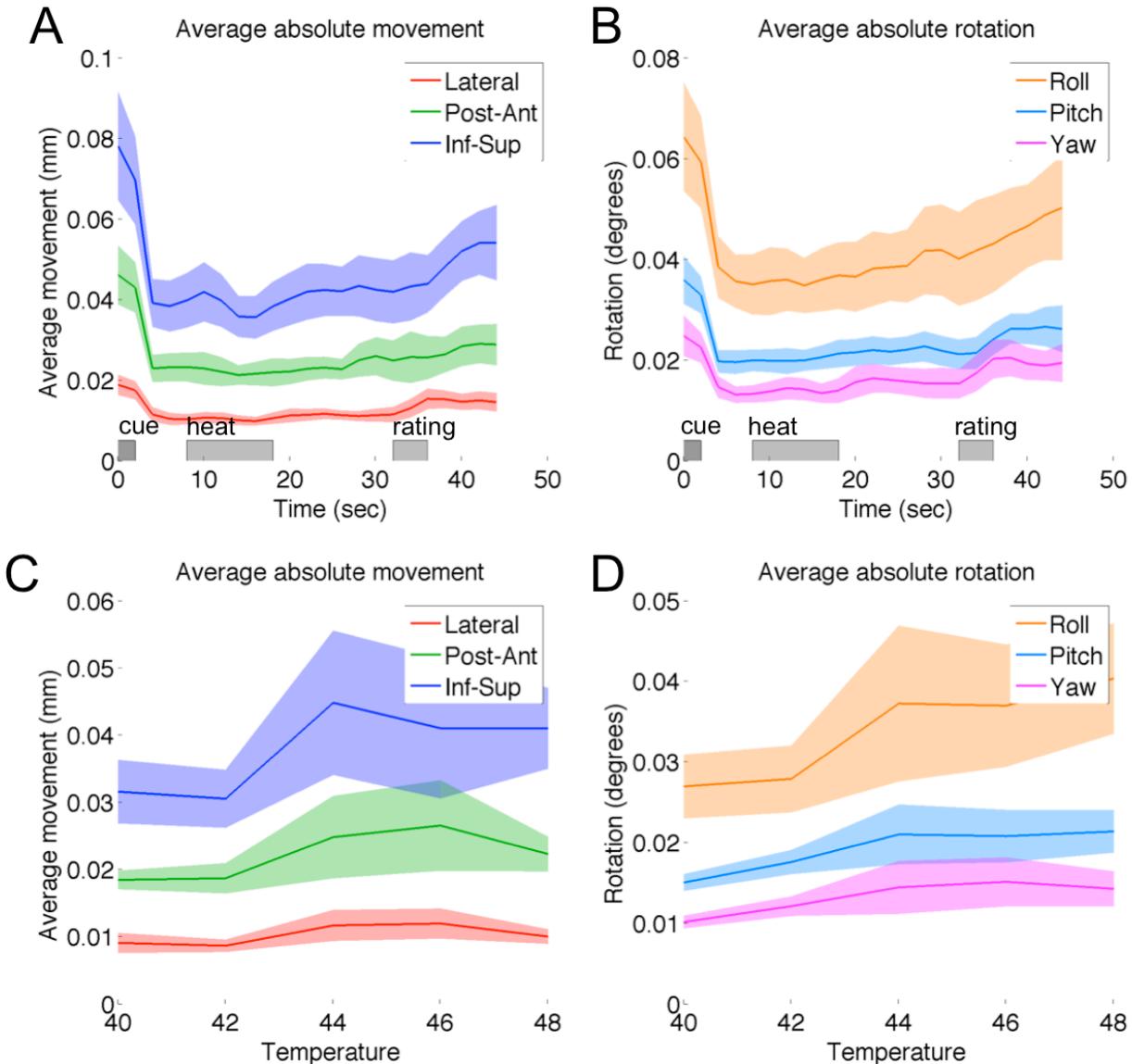
The time course of signature response

To examine the time course of the signature response during thermal stimulation and further assess the relationship with pain vs. heat sensation across time, we reconstructed signature response every 2 sec during the various phases of the stimulation trials: anticipation of pain, pain experience, pain judgment, and rest (Supplementary Figure S4). Signature response rose during the application of heat and monotonically tracked the actual temperatures, but did not respond to anticipatory cues or post-pain decision-making periods, demonstrating specificity to the time period when pain was experienced. In addition, stimulus delivery and subjective pain follow different time courses due to temporal summation^{16,26}, permitting a test of which correlates more highly with signature response. We estimated the time course of subjective pain during heat epochs in a separate sample ($N = 12$), and convolved that time course with the canonical SPM hemodynamic response function to obtain a prediction based on expected moment-by-moment pain experience (purple in Fig. S4B). We contrasted that with a model in which the time course of stimulation itself was convolved with the canonical SPM hemodynamic response function to obtain a prediction based on moment-by-moment heat intensity.

We estimated the slope of the relationship between signature activity and temperature at each time point for each participant. Correlation between the time course of signature temperature effects (slopes) and predicted fMRI responses were higher for the pain report

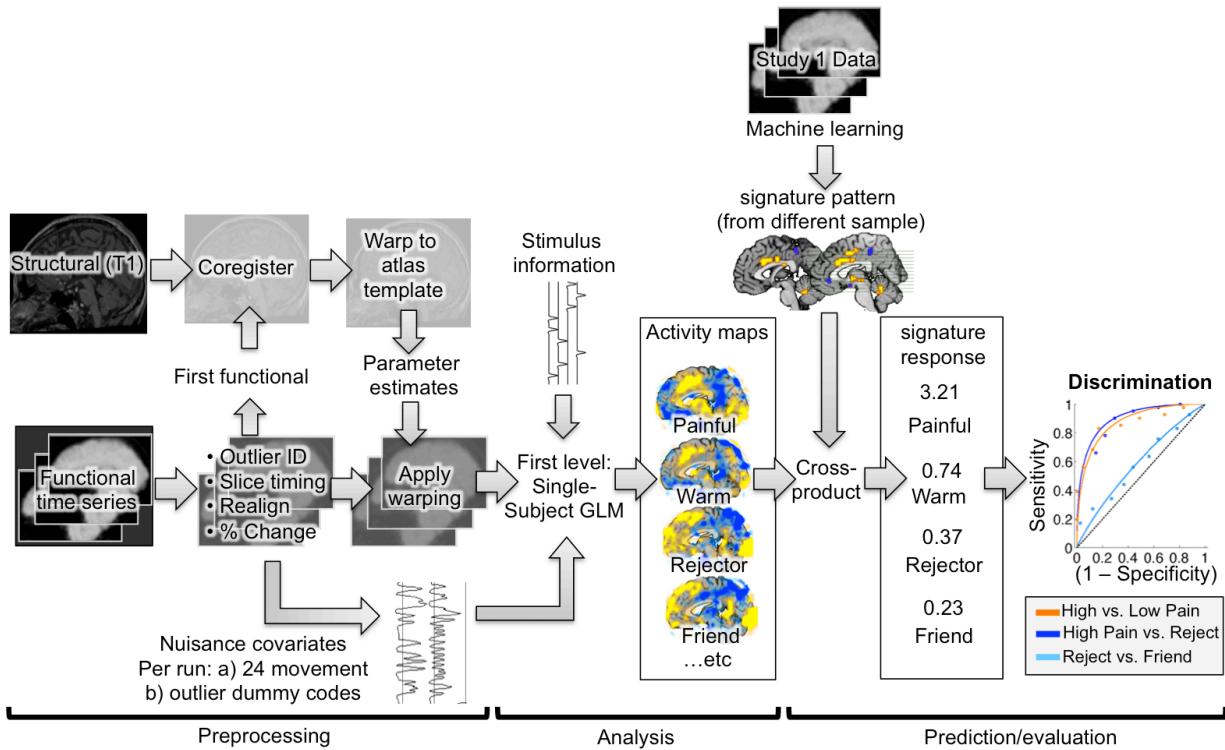
predictor than the stimulation time course for every individual tested ($r = 0.89 \pm 0.007$ vs. $r = 0.76 \pm 0.01$, respectively; $P < 0.001$; Fig. S4C). These results further suggest specificity to pain experience rather than general salience, somatic sensation, or decision processes.

Supplementary Figure S1



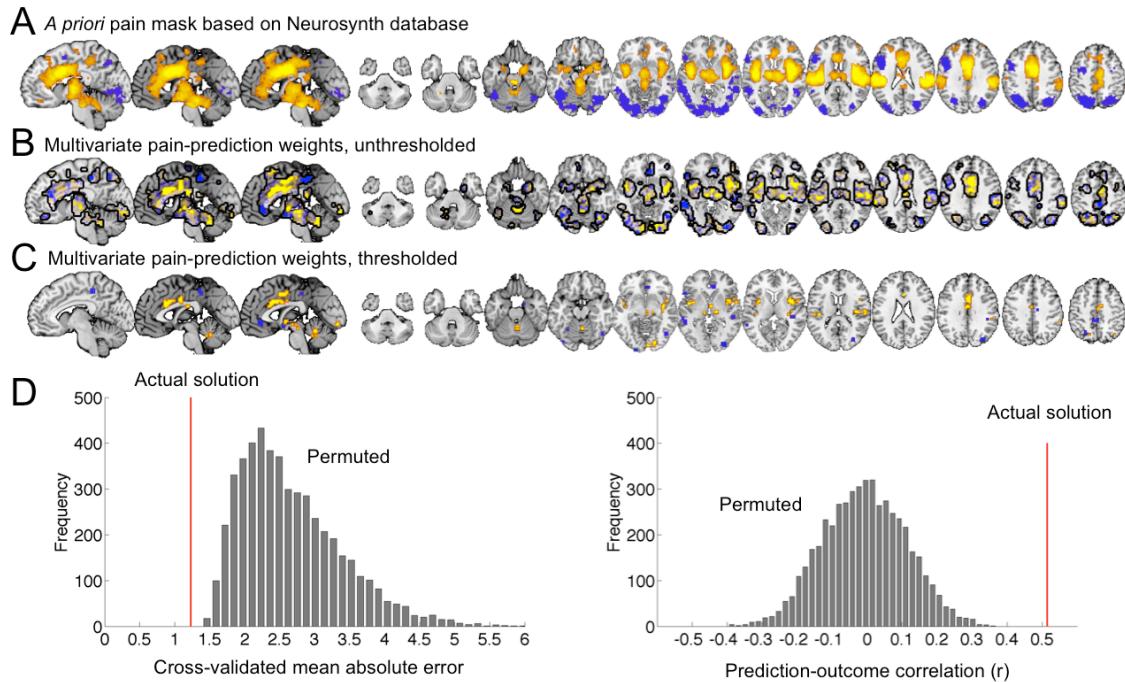
Head movement in Study 1. Three translation (A, C) and three rotation (B, D) parameter estimates, based on image realignment, are plotted as a function of time within the heat trial (A, B) and stimulus temperature (C, D). In each case, the average absolute displacement from the previous image is plotted on the y-axis. Error bars show standard error of the mean. Head movement did not increase during stimulation or at stimulus onset and offset. Rather, a modest movement increase is observed at the onset of the pain-predictive cue. Movement was not significantly predicted by temperature for any movement direction.

Supplementary Figure S2



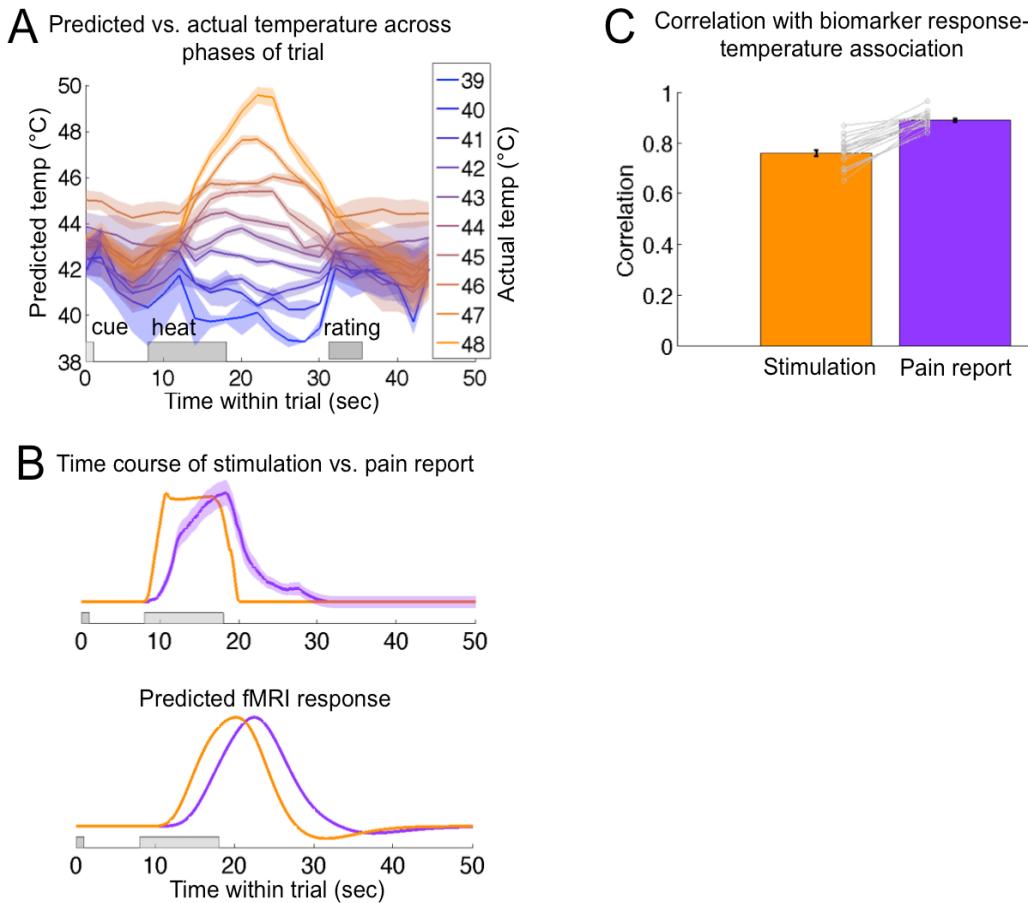
Preprocessing and analysis stages. The preprocessing and first-level General Linear Model (GLM) are standard steps performed with SPM software, with the exception of outlier identification and percent-change scaling. Activity maps from the GLM are cross-multiplied by the signature map, which was developed using a separate cross-validated machine learning regression (not illustrated), to yield a scalar signature response value for each image. Signature response values are used to predict continuous pain and in classification.

Supplementary Figure S3



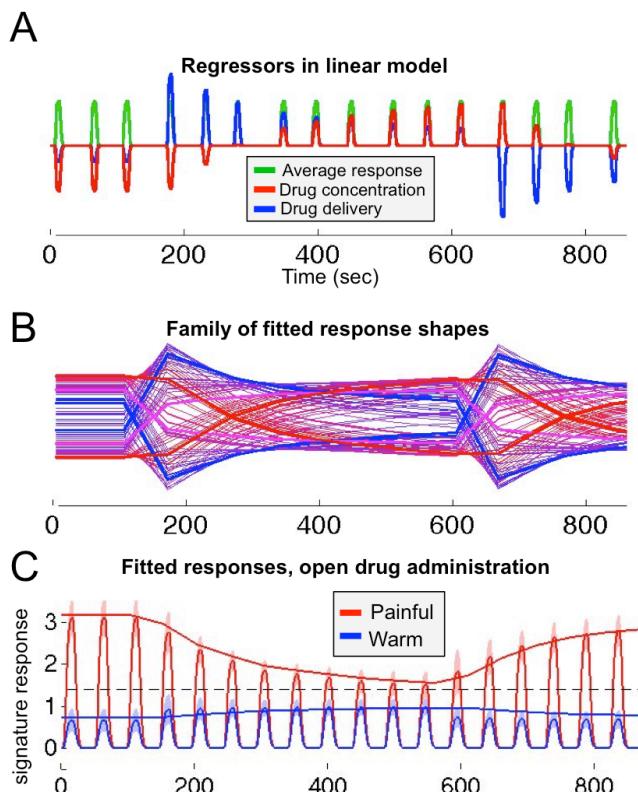
Signature development in Study 1. A) A mask of *a priori* regions used in analysis based on the Neurosynth database, associated with ‘pain’ at $q < 0.05$ FDR-corrected. In all plots, yellow indicates positive predictive weights for pain, and blue indicates negative weights. B) Unthresholded signature pattern weights from the LASSO-PCR analysis, shown as Z-scores, with voxels with lower Z-scores more transparent. The black outline shows the *a priori* mask boundaries. Blue/yellow indicate $Z < -2$ and $Z > 2$, respectively. C) Map thresholded at $q < 0.05$ FDR ($P < 0.003$) for display. Blue/yellow indicate $Z < -3$ and $Z > 3$, respectively. D) Histograms of prediction error and prediction-outcome correlation from nonparametric permutation test. Histograms show the distribution of null-hypothesis results, and the red line shows the actual solution.

Supplementary Figure S4



Correlation of signature response with the time course of objective stimulus delivery vs. reported pain in Study 1. A) Signature response (scaled to reflect predicted temperature) across time within trials. Lines/shading: means/standard errors across participants. Pattern expression increased monotonically with temperature only following stimulation, and not during cue and pain report periods. B) Top: Time-course of thermal stimulation (orange) and subjective pain (purple; shaded area: SEM). Bottom: Predicted fMRI activity, convolving the stimulus and report time-courses with SPM's standard double-gamma hemodynamic response function. The predictors were correlated ($r = 0.78$, 61% of variance shared), but the pain time course peaked appreciably later. C) Correlation between the time course of signature temperature effects and the model were higher for the pain report model (purple) than the stimulation time course model (orange) for every individual tested. Correlations for individual subjects are shown by points connected with light gray lines.

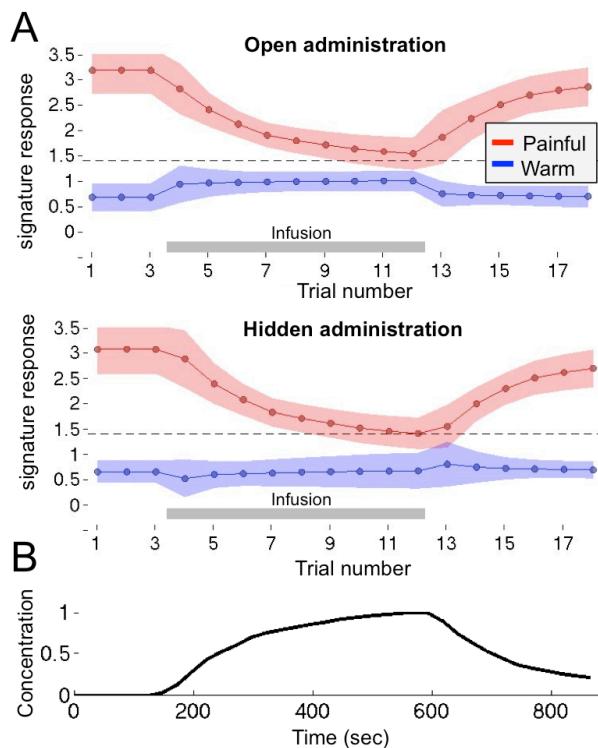
Supplementary Figure S5



Modeling of drug effects in Study 4. A) Regressors for the average response (green), drug effect site concentration estimated using a pharmacokinetic model (red), and the drug infusion period itself, orthogonalized to the drug regressor (blue). B) The family of shapes modeled by the modulator regressors (red and blue). C) The average fitted signature response for hot and warm trials in the Open administration condition. Responses in the Hidden condition were not discernably (or statistically) different. Shaded areas show the standard error of the mean across participants. The solid lines show the fitted responses across trial amplitude estimates. The horizontal dashed line shows the threshold for Hot vs. Warm trials from Study 1.

Supplementary Figure S6.

Figure S6. signature response to remifentanil in Study 4



Signature response to remifentanil in Study 4. A) The signature from Study 1 applied to Painful (red) and Warm (blue) events across trials. The gray box marks the intravenous drug infusion period. Average model fits with SEM across individuals (shaded areas) are shown. The model captured the effects of drug effect site concentration and the infusion period itself on responses to Painful and Warm events; thus, the curves reflect a combination of potential drug and psychological effects across time. B) Average profile of drug effect site concentration based on the pharmacokinetic model of Minto et al²⁴. The observed signature responses parallel the time course of effect site concentration and show no effect of Open vs. Hidden administration. Both findings suggest that signature responses are mainly influenced by the drug itself, rather than expectations about drug delivery.

Supplementary Table S1

Name	x	y	z	mm ³	Z	Name	x	y	z	mm ³	Z
Thermal pain: Positive predictive weights											
Vermis (CBLM)	2	-53	-20	486	3.35	R ITC	47	-62	-8	432	-3.35
R Ant/MidINS	38	4	4	2241	3.35	L Fusiform gyrus	-40	-56	-17	81	-3.35
L Superior temporal gyrus	-40	-11	-8	162	3.35	L Inferior Occipital gyrus	-40	-80	-11	378	-3.35
R Calcarine gyrus (BA17)	8	-89	-5	189	3.35	L Inferior Occipital gyrus	-34	-65	-8	162	-3.35
R vIThAl	14	-17	1	405	3.35	R gyrus (BA18)	-22	-98	-5	81	-3.35
L midINS	-37	4	4	810	3.35	vmPFC	8	37	1	405	-3.35
Hypothal	2	-5	1	216	3.35	L Middle temporal gyrus	-55	-41	4	567	-3.35
L vIThAl	-13	-17	1	81	3.04	L IFG	-52	25	4	162	-3.35
R frOP / temporal pole	59	4	1	189	3.35	R Inferior Occipital gyrus	38	-83	4	81	-3.16
L dpIns/SII	-40	-20	13	270	3.35	R Heschl's Gyrus	41	-26	10	162	-3.35
R dpINS	41	-17	13	324	3.35	R Middle Occipital Gyrus	32	-77	19	216	-3.35
R SII	59	-17	16	162	3.04	R PCC/Precuneus/paracentral lobule	32	-77	34	270	-3.35
L TPJ (Superior temporal gyrus)	-64	-32	22	216	3.35	R SPL	-1	-35	49	513	-3.35
dACC	2	13	31	1917	3.35	L SPL	23	-62	55	297	-3.35
R Supramarginal gyrus	53	-32	31	108	3.35	R Middle Occipital Gyrus	-19	-65	61	189	-3.35
R IPL	59	-35	37	162	3.16	Gyrus	35	-89	4	513	3.35

Note. Peak coordinates from the machine learning analysis in Study 1. The signature map was thresholded at $q < 0.05$ FDR for interpretation, based on a bootstrap test with 5000 bootstrap samples. Peak coordinates for positive and negative weights are listed in the left and right columns, respectively. Coordinates are reported in standard Montreal Neurologic Institute space. ACC, anterior cingulate cortex; CBLM: cerebellum; IFG, inferior frontal gyrus; INS, insula; IPL, inferior parietal lobule; ITC, inferior temporal cortex; OCC, occipital; frOP, frontal operculum; PCC, posterior cingulate cortex; PHCMP, parahippocampal cortex; PFC, prefrontal cortex; SMA, supplementary motor cortex; SPL, superior parietal lobule; STS, superior temporal sulcus; Thal, thalamus; TPJ, temporal-parietal junction; mvPFC, ventromedial prefrontal cortex. Prefixes: a, anterior; d, dorsal; l, lateral; m, medial; r, rostral; s, superior; v, ventral.

Supplementary Table S2.

Forced-choice discrimination test	Discrimination Sens./Spec./PPV^h	Effect size AUC	Binomial test d_a	P-value
Study 1				
Painful vs. Warm ^a	100% (100-100%)	1.00	4.88	P < 0.001
Pain vs. Anticipation	100% (100-100%)	1.00	3.92	P < 0.001
Pain vs. Pain Recall	100% (100-100%)	1.00	2.29	P < 0.001
Conditions different by 3+ VAS units ^f	100% (100-100%)	1.00	3.91	P < 0.001
Conditions different by 2-3 VAS units	93% (84-100%)	0.97	2.17	P < 0.001
Conditions different by 1-2 VAS units	86% (76-95%)	0.86	1.15	P < 0.001
Conditions different by 0.5-1 VAS unit	69% (50-90%)	0.80	0.99	P = 0.26
Study 2				
Painful vs. Warm ^c	100% (100-100%)	1.00	3.12	P < 0.001
Painful (>125) vs. near-threshold (75-125) ^e	100% (100-100%)	1.00	2.77	P < 0.001
High (50-100) vs. low (0-50) warmth	100% (100-100%)	1.00	2.18	P < 0.001
49.3 ^g vs. 48.3°C	90% (81%-97%)	0.93	1.71	P < 0.001
48.3 vs. 47.3°C	100% (100-100%)	1.00	2.00	P < 0.001
47.3 vs. 46.3°C	80% (67%-91%)	0.82	0.96	P = 0.001
46.3 vs. 45.3°C	67% (53%-81%)	0.77	0.77	P = 0.10
45.3 vs. 44.3°C	70% (56%-83%)	0.66	0.43	P = 0.04
Study 3				
Painful vs. Warm	93% (86-98%)	0.97	2.08	P < 0.001
Painful vs. Rejector Photo	95% (89-100%)	0.98	2.09	P < 0.001
Rejector Photo vs. Friend Photo	56% (43-69%)	0.66	0.49	P = 0.53
Study 4				
Hot vs. Warm, pre-drug	90% (79-100%)	0.97	1.76	P < 0.001
Hot vs. Warm, on-drug	76% (61-90%)	0.84	1.08	P < 0.05
Hot pre-drug vs. on-drug	76% (60-92%)	0.84	1.08	P < 0.05

Note. Forced-choice classification performance across studies. *a*: Painful conditions were defined as those > 44.5° C and > 5.80 average VAS units, and Warm as < 44.5° C and < 3.34 VAS units. *b*: Study 2 was conducted on a scanner with a different field strength (3T), so a new threshold was estimated. *c*: Participants made painful vs. non-painful judgments on each trial. *d*: The threshold derived from Study 1 was applied. *e*: Continuous, 100-point VAS ratings for pain or warmth intensity (0-99 for warmth, 100-200 for pain). *f*: Visual analogue scale (VAS) ratings on a continuous, 9-point scale. *g*: Only 4 trials were included at 49.3° (cf. 11 trials for 44.3° and 15 trials for other conditions.) *h*: For two-choice (forced-choice) discrimination, the decision threshold (for the difference between pairs) is 0, and the sensitivity, specificity, and positive predictive value (PPV) are the same, and are equal to the decision accuracy. AUC: Area under the Receiver Operating Characteristic curve, a threshold-independent measure of performance; chance is 0.5. PPV: Positive predictive value. d_a: Discriminability, a measure

of effect size under a Gaussian model. Performance varies to some degree based on the number of trials per subject averaged to form condition maps in each study.

Supplementary Table S3

<u>Pain/no pain test</u>	Thresh	Sensitivity	Specificity	PPV	AUC	d _a	Effect size	Binomial P-value
Anterior insula								
Painful vs. Warm	0.44	83% (73-92%)	83% (73-92%)	83% (73-92%)	0.85	1.46	P < 0.001	
Painful vs. Rejector Photo	0.48	80% (70-91%)	80% (69-90%)	80% (70-90%)	0.81	1.32	P < 0.001	
Rejector Photo vs. Friend Photo	0.28	39% (26-50%)	78% (66-88%)	64% (47-79%)	0.53	0.15	P = 0.15	
Anterior cingulate								
Painful vs. Warm	0.22	61% (49-74%)	85% (56-94%)	81% (69-92%)	0.75	1.02	P < 0.001	
Painful vs. Rejector Photo	0.07	73% (62-84%)	76% (65-86%)	75% (64-86%)	0.78	1.15	P < 0.001	
Rejector Photo vs. Friend Photo	0.12	22% (12-33%)	90% (82-98%)	69% (46-90%)	0.53	0.21	P = 0.32	
S2/dorsal posterior insula								
Painful vs. Warm	0.19	73% (62-85%)	54% (40-67%)	61% (50-73%)	0.63	0.57	P = 0.02	
Painful vs. Rejector Photo	0.12	73% (61-84%)	88% (78-96%)	86% (75-95%)	0.86	1.57	P < 0.001	
Rejector Photo vs. Friend Photo	-0.58	59% (45-71%)	59% (46-71%)	59% (45-72%)	0.55	0.18	P = 0.15	
<u>Forced-choice test</u>		Discrimination			Effect size		Binomial	
		Sens./Spec./PPV^a			AUC		P-value	
Anterior insula								
Painful vs. Warm		88% (79-95%)			0.95	1.57	P < 0.001	
Painful vs. Rejector Photo		83% (73-92%)			0.91	1.39	P < 0.001	
Rejector Photo vs. Friend Photo		56% (44-68%)			0.59	0.25	P = 0.53	
Anterior cingulate								
Painful vs. Warm		80% (70-91%)			0.86	1.12	P < 0.001	
Painful vs. Rejector Photo		80% (70-90%)			0.89	1.20	P < 0.001	
Rejector Photo vs. Friend Photo		61% (48-73%)			0.62	0.34	P = 0.21	
S2/dorsal posterior insula								
Painful vs. Warm		73% (61-84%)			0.73	0.69	P < 0.01	
Painful vs. Rejector Photo		90% (82-98%)			0.97	1.71	P < 0.001	
Rejector Photo vs. Friend Photo		59% (46-71%)			0.61	0.31	P = 0.35	

Note. Pain/no-pain classification for selected single regions in Study 3. The signature response estimated from weights only with each single, a priori region of interest were used to evaluate performance. *a*: For two-choice (forced-choice) discrimination, the decision threshold (for the difference between pairs) is 0, and the sensitivity, specificity, and positive predictive value (PPV) are the same, and are equal to the decision accuracy. AUC: Area under the Receiver Operating Characteristic curve, a threshold-independent measure of performance. PPV: Positive predictive value. d_a : Discriminability, a measure of effect size under a Gaussian model. 95% confidence intervals are shown in parentheses, and were estimated using bootstrap resampling with 1000 samples per test.

References

1. Atlas LY, Bolger N, Lindquist MA, Wager TD. Brain Mediators of Predictive Cue Effects on Perceived Pain. *J Neurosci* 2010;30:12964–77.
2. Buhle J, Wager TD. Performance-dependent inhibition of pain by an executive working memory task. *Pain* 2010.
3. Ashburner J, Friston KJ. Unified segmentation. *NeuroImage* 2005;26:839-51.
4. Lund TE, Madsen KH, Sidaros K, Luo WL, Nichols TE. Non-white noise in fMRI: does modelling have an impact? *NeuroImage* 2006;29:54-66.
5. Handwerker DA, Gonzalez-Castillo J, D'Esposito M, Bandettini PA. The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *NeuroImage* 2012.
6. Aguirre G, Zarahn E, D'Esposito M. The variability of human, BOLD hemodynamic responses. *NeuroImage* 1998;8:360-9.
7. Thomas CG, Harshman RA, Menon RS. Noise reduction in BOLD-based fMRI using component analysis. *NeuroImage* 2002;17:1521-37.
8. Tohka J, Foerde K, Aron AR, Tom SM, Toga AW, Poldrack RA. Automatic independent component labeling for artifact removal in fMRI. *NeuroImage* 2008;39:1227-45.
9. Tibshirani R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B* 1996;58:267-88.
10. Wager TD, Atlas LY, Leotti LA, Rilling JK. Predicting Individual Differences in Placebo Analgesia: Contributions of Brain Activity during Anticipation and Pain Experience. *J Neurosci* 2011;31:439-52.
11. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods* 2011.
12. Baliki MN, Geha PY, Apkarian AV. Parsing pain perception between nociceptive representation and magnitude estimation. *J Neurophysiol* 2009;101:875-87.
13. Lindquist MA, Loh JM, Atlas L, Wager TD. Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. *NeuroImage* 2009;45:S187-S98.
14. Wager TD, Rilling JK, Smith EE, et al. Placebo-induced changes in FMRI in the anticipation and experience of pain. *Science* 2004;303:1162-7.
15. Bornhovd K, Quante M, Glauche V, Bromm B, Weiller C, Buchel C. Painful stimuli evoke different stimulus-response functions in the amygdala, prefrontal, insula and somatosensory cortex: a single-trial fMRI study. *Brain* 2002;125:1326-36.
16. Koyama Y, Koyama T, Kroncke AP, Coghill RC. Effects of stimulus duration on heat induced pain: the relationship between real-time and post-stimulus pain ratings. *Pain* 2004;107:256-66.
17. Rish I, Cecchi G, Baliki M. Sparse regression models of pain perception. *Brain Informatics* 2010.
18. Grosenick L, Greer S, Knutson B. Interpretable classifiers for FMRI improve prediction of purchases. *IEEE Trans Neural Syst Rehabil Eng* 2008;16:539-48.
19. Smola AJ, Schölkopf B. A tutorial on support vector regression. *Statistics and computing* 2004;14:199-222.
20. Kross E, Berman MG, Mischel W, Smith EE, Wager TD. Social rejection shares somatosensory representations with physical pain. *Proceedings of the National Academy of Sciences* 2011;108:6270-5.
21. Kross E, Davidson M, Weber J, Ochsner K. Coping with emotions past: the neural bases of regulating affect associated with negative autobiographical memories. *Biol Psychiatry* 2009;65:361-6.
22. Wager TD, Scott DJ, Zubieta JK. Placebo effects on human mu-opioid activity during pain. *Proceedings of the National Academy of Sciences* 2007;104:11056-61.
23. Atlas LY, Whittington RA, Lindquist MA, Wielgosz J, Sonty N, Wager TD. Dissociable influences of opiates and expectations on pain. *J Neurosci* 2012;32:8053-64.
24. Minto CF, Schnider TW, Egan TD, et al. Influence of age and gender on the pharmacokinetics and pharmacodynamics of remifentanil. I. Model development. *Anesthesiology* 1997;86:10-23.
25. Minto CF, Schnider TW, Shafer S. Pharmacokinetics and pharmacodynamics of remifentanil. II. Model application. *Anesthesiology* 1997;86:24-33.
26. Apkarian AV, Darbar A, Krauss BR, Gelnar PA, Szeverenyi NM. Differentiating cortical areas related to pain perception from stimulus identification: temporal analysis of fMRI activity. *J Neurophysiol* 1999;81:2956.
27. Wanigasekera V, Lee MC, Rogers R, Hu P, Tracey I. Neural correlates of an injury-free model of central sensitization induced by opioid withdrawal in humans. *J Neurosci* 2011;31:2835-42.

28. Wise RG, Williams P, Tracey I. Using fMRI to quantify the time dependence of remifentanil analgesia in the human brain. *Neuropsychopharmacology* 2004;29:626-35.
29. Buhle JT, Stevens BL, Friedman JJ, Wager TD. Distraction and Placebo: Two Separate Routes to Pain Control. *Psychol Sci* 2012.
30. Bantick SJ, Wise RG, Ploghaus A, Clare S, Smith SM, Tracey I. Imaging how attention modulates pain in humans using functional MRI. *Brain* 2002;125:310-9.
31. Apkarian AV, Hashmi JA, Baliki MN. Pain and the brain: Specificity and plasticity of the brain in clinical chronic pain. *Pain* 2011;152:S49-S64.
32. Borsook D, Becerra L, Hargreaves R. Biomarkers for chronic pain and analgesia. Part 1: the need, reality, challenges, and solutions. *Discovery medicine* 2011;11:197-207.
33. Borsook D, Becerra L, Hargreaves R. Biomarkers for chronic pain and analgesia. Part 2: how, where, and what to look for using functional imaging. *Discovery medicine* 2011;11:209-19.
34. McGonigle DJ, Howseman AM, Athwal BS, Friston KJ, Frackowiak RS, Holmes AP. Variability in fMRI: an examination of intersession differences. *NeuroImage* 2000;11:708-34.
35. Glover GH, Mueller BA, Turner JA, et al. Function biomedical informatics research network recommendations for prospective multicenter functional MRI studies. *J Magn Reson Imaging* 2012.
36. Hoge RD. Calibrated fMRI. *NeuroImage* 2012.
37. Björnsdotter M, Loken L, Olausson H, Vallbo A, Wessberg J. Somatotopic organization of gentle touch processing in the posterior insular cortex. *J Neurosci* 2009;29:9314-20.
38. Haynes JD, Rees G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 2005;8:686-91.
39. Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 2005;8:679-85.
40. Sabuncu MR, Singer BD, Conroy B, Bryan RE, Ramadge PJ, Haxby JV. Function-based intersubject alignment of human cortical anatomy. *Cereb Cortex* 2010;20:130-40.
41. Flor H, Elbert T, Knecht S, et al. Phantom-limb pain as a perceptual correlate of cortical reorganization following arm amputation. *Nature* 1995;375:482-4.
42. Donahue MJ, Hoogduin H, van Zijl PC, Jezzard P, Luijten PR, Hendrikse J. Blood oxygenation level-dependent (BOLD) total and extravascular signal changes and DeltaR2* in human visual cortex at 1.5, 3.0 and 7.0 T. *NMR Biomed* 2011;24:25-34.
43. Davis TL, Kwong KK, Weisskoff RM, Rosen BR. Calibrated functional MRI: mapping the dynamics of oxidative metabolism. *Proc Natl Acad Sci U S A* 1998;95:1834-9.