nature research

https://doi.org/10.1038/s41591-020-1142-7

Supplementary information

A neuroimaging biomarker for sustained experimental and clinical pain

In the format provided by the authors and unedited

SUPPLEMENTARY INFORMATION:

A neuroimaging biomarker for sustained experimental and clinical pain

Jae-Joong Lee[1,2], Hong Ji Kim[1,2], Marta Čeko[3,4], Bo-yong Park[1,5], Soo Ahn Lee[1,2], Hyunjin Park[1,6], Mathieu Roy[7,8], Seong-Gi Kim[1,2], Tor D. Wager[9]*, Choong-Wan Woo[1,2,10,11]*

[1]Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, South Korea

[2]Department of Biomedical Engineering, Sungkyunkwan University, Suwon, South Korea

[3]Institute of Cognitive Science, University of Colorado, Boulder, USA

[4]Department of Psychology and Neuroscience, University of Colorado, Boulder, USA

[5]McConnell Brain Imaging Centre, Montreal Neurological institute and Hospital, McGill University, Montreal, QC, Canada.

[6]School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, South Korea

[7]Department of Psychology, McGill University, Montreal, QC, Canada

[8]Alan Edwards Centre for Research on Pain, McGill University, Montreal, QC, Canada

[9]Department of Psychological and Brain Sciences, Dartmouth College

[10]Biomedical Institute for Convergence at SKKU, Sungkyunkwan University, Suwon, South Korea

[11]Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Suwon, Korea

*Co-corresponding authors

This PDF file includes:

1. Supplementary Methods
2. Supplementary Results
3. Supplementary Figures 1-9
4. Supplementary Tables 1-4

SUPPLEMENTARY METHODS

Study 1: Capsaicin-induced tonic pain dataset (training dataset)

Participants. Nineteen healthy and right-handed participants were included (age = 23.2 ± 4.9 [mean ± SD], 10 female). Participants were recruited from Boulder/Denver Metro Areas. The institutional review board of the University of Colorado Boulder approved the study. All participants provided written informed consent. Preliminary eligibility of participants was determined through an online questionnaire. Participants with psychiatric, neurological, or systemic disorders and MRI contraindications were excluded.

Experimental design. Study 1 had two condition runs; one was the capsaicin run, and the other was the control run. In the capsaicin run, we applied hot chili pepper sauce on participants' tongue (for more details, see 'capsaicin delivery' above). The control run did not include any external stimulation. These two runs are conducted consecutively, and structural scan was obtained between them to minimize any remaining sensations of the preceding run. The order of capsaicin and control runs was counterbalanced across participants. Each run lasted for 5 minutes 15 seconds, and participants provided ratings of pain intensity and unpleasantness every 45 seconds from the start of the scan (total 7 times) using a MR-compatible trackball device. After the scan, we asked multiple post-scan questions to participants regarding their thought contents during the scan. These data were not used in the current study. To minimize the unwanted effect of remaining painful sensation after the capsaicin run, we provided a small amount of liquid sucrose after the scan until the participant reported no residual pain. Graphical illustration of the experimental design of Study 1 is provided in Fig. 1b.

Rating scale. The general Labeled Magnitude Scale (gLMS)[1] was used for ratings. The anchors of the gLMS began with "Not at all (0)" to the far left of the scale, and continued to the right in a graded fashion with anchors of "A little bit" (0.061), "Moderately" (0.172), "Strongly" (0.354), and "Very strongly" (0.533), until "Most (Strongest imaginable sensation/unpleasantness of any kind)" (1) on the far right.

Physiological data acquisition. Physiological responses including skin conductance response (SCR) and cardiac signal were obtained during the scan (Biopac systems, Goleta, CA). SCR was recorded using Ag/Agcl electrodes placed on the medial phalanges of the middle and ring fingers of the left hand. Cardiac signal was recorded using Photo Plethysmogram (PPG) amplifier placed on the big toe. Both SCR and PPG data were sampled at 2000Hz. We note that physiological data of five participants were discarded, because they were not recorded during scan or the data quality was too bad.

Physiological data analysis. For PPG data, inter-beat intervals (IBI) were calculated by finding peaks and getting distance between them. IBI data and SCR data were low-pass filtered (5 Hz) and down-sampled (25 Hz). Heart rate (HR) was calculated by taking the inverse of IBI data (1/IBI). For comparison with pain rating, we divided the HR and SCR timeseries into 7 bins that corresponded to each intermittent self-report of pain rating (total of 7 ratings). Each bin was defined as an interval between the report of one pain rating and that of next (immediately after) pain rating. Because the last pain rating had no 'next' report of pain by its definition, the final bin ended at the end of the scan. Each bin of HR and SCR signals were then averaged.

fMRI data acquisition. Whole-brain fMRI data were acquired on a 3T Siemens TrioTim scanner at the University of Colorado Boulder. High-resolution T1-weighted structural images

were acquired. Functional EPI images were acquired with TR = 460 ms, TE = 29.0 ms, multiband acceleration factor = 8, field of view = 248 mm, 82×82 matrix, 3×3×3 mm^3 voxels, 56 interleaved slices, number of volumes = 685. Stimulus presentation and behavioral data acquisition were controlled using E-Prime software (PST Inc.).

fMRI data analysis. Structural and functional MRI data were preprocessed using an automated preprocessing pipeline, which was based on AFNI, FSL, and SPM5 and developed at Mind Research Network (MRN)[2]. In the pipeline, structural T1-weighted images were co-registered to the functional image and normalized to MNI, and functional EPI images were distortion-corrected, motion-corrected (realigned), normalized to MNI using T1 images with the interpolation to 3×3×3 mm^3 voxels, and smoothed with a 6 mm FWHM Gaussian kernel. After the automated preprocessing, 20 initial volumes of fMRI data were removed in order to allow for image intensity stabilization. Then, we concatenated the fMRI data across the capsaicin and control runs before further preprocessing (e.g., nuisance regression). We did the concatenation in order to preserve the difference between the capsaicin and control conditions within an individual because we used the combined data across two conditions for model building given that the capsaicin run of Study 1 was not long enough to capture the rise and fall of the painful sensation (for this reason, we used 20-minute scans in Study 3). Conducting further preprocessing separately on each run could obscure the differences between two conditions. After the concatenation, we regressed out the nuisance covariates, which included (i) image-intensity outliers (i.e., "spikes"), (ii) finger movement period indicators related with report of pain rating, and (iii) 24 head motion parameters (6 movement parameters including x, y, z, roll, pitch, and yaw, their mean- centered squares, their derivatives, and squared derivative)[3]. Outliers were identified based on mean signal intensity, mahalanobis distances, and the root mean square of successive difference across volumes. Images that exceeds 3 standard deviation from the global mean were considered as outliers. For signal intensity and mahalanobis distances, the images that exceeded 10 mean absolute deviations (MADs) based on moving averages with full width at half maximum (FWHM) 20 images kernel were additionally identified as outliers. Each time-point identified as outliers by either outlier detection method was included as nuisance covariates. After the nuisance regression, intensity value outside of the median $\pm$ 5 standard deviation of all intensity values (across all the voxels and time points) were winsorized, and low pass filter with 0.1 Hz was applied.

Study 2: Capsaicin-induced tonic pain dataset (validation dataset)

Participants. Forty-two healthy and right-handed participants were included (age = 24.4 ± 6.0 [mean ± SD], 14 female) after excluding 7 participants who provided avoidance rating scores higher in the control run than the capsaicin run from further analysis. Other information is same with Study 1.

Experimental design. Study 2 had three conditions; (1) capsaicin, (2) bitter taste (quinine), and (3) control condition runs. For the capsaicin and bitter taste delivery procedure, please see 'capsaicin delivery' and 'aversive tastant delivery' above. The three runs and structural scan were conducted consecutively, the order of which was counterbalanced across participants. To collect ratings on a common scale across different stimulus modalities, we used an avoidance rating scale instead of pain intensity or unpleasantness. The question prompt was "How much do you want to avoid this experience in the future?" Each run lasted for 5 minutes

10 seconds, and participants provided ratings of avoidance every 30 seconds starting from 10 seconds after the scan started (total 10 times) using a trackball. Other procedures were same to Study 1 including post-run questions and delivering liquid sucrose between runs.

Rating scale. The general Labeled Magnitude Scale (gLMS)[1] was used for the avoidance rating. The anchors of the gLMS began with "Not at all (0)" to the far left of the scale, and continued to the right in a graded fashion with anchors of "A little bit" (0.061), "Moderately" (0.172), "Strongly" (0.354), and "Very strongly" (0.533), until "Most (I never want to experience this again in my life)" (1) on the far right.

fMRI data acquisition. Whole-brain fMRI data were acquired on a 3T Siemens Prisma scanner at the University of Colorado Boulder. High-resolution T1-weighted structural images were acquired. Functional EPI images were acquired with TR = 460 ms, TE = 27.2 ms, multiband acceleration factor = 8, field of view = 220 mm, 82×82 matrix, 2.7×2.7×2.7 mm^3 voxels, 56 interleaved slices, number of volumes = 676. Stimulus presentation and behavioral data acquisition were controlled using Matlab (Mathworks) and Psychtoolbox (http://psychtoolbox.org/).

fMRI data analysis. Preprocessing was done with the same pipeline as Study 1.

Study 3: Capsaicin-induced tonic pain dataset (independent test dataset)

Participants. Forty-eight healthy, right-handed participants were included (age = 22.8 ± 2.4 [mean ± SD], 21 female) after we excluded 4 participants who provided avoidance rating scores higher in the control run than the capsaicin run and one participant whose brain coverage of MRI was insufficient to contain the whole brain. Participants were recruited from Suwon area in South Korea. The institutional review board of the Sungkyunkwan University approved the study. Other information is same with Studies 1 and 2.

Experimental design. Study 3 had four conditions total; (i) capsaicin, (ii) bitter taste (quinine), (iii) aversive odor (fermented skate), and (iv) control runs. For the capsaicin, bitter taste, aversive odor delivery procedure, please see 'capsaicin delivery', 'aversive tastant delivery', and 'aversive odor delivery' above. After a structural scan, the four condition runs were administered consecutively, the order of which was counterbalanced across participants. Each run lasted for 20 minutes, and participants provided avoidance ratings continuously throughout the run using a trackball. We designed the experiment to have long scans to capture the full rise and fall of each sensation. To prevent participants falling asleep and help maintain a certain level of alertness during the scan, we used an intermittent simple response task, in which the color of the rating bar on the screen was changed from orange to red for 1 second every minute with some jitter, and the participants had to respond to the color change by clicking the button of the trackball device. Other procedures were same to Study 2 including post-run questions and delivering liquid sucrose if participants reported that there was a remaining taste sensation between runs.

Physiological data acquisition. Physiological responses including cardiac and respiratory signals were obtained during the scan (Biopac systems, Goleta, CA). The cardiac signal was recorded using Electrocardiogram (ECG) amplifiers placed on the right clavicle and left lower rib area. The respiratory signal was recorded using a pneumatic belt placed around the

chest. We note that physiological data of ten participants had to be discarded (remaining 38 participants) because of their poor data quality. The sampling rate was 5000 Hz for all physiological data.

Physiological data analysis. To reduce MR-related artifact from physiological data, ECG data was preprocessed with a band-stop filter that removes fundamental and harmonics of MRI sampling frequency (here, multiplies of 1/TR = 2.1739 Hz) and low-pass filter (< 5 Hz). R peaks were detected in this denoised ECG signal using the PhysIO toolbox (obtained from https://github.com/translationalneuromodeling/tapas/tree/master/PhysIO). Inter-beat interval was calculated by getting distance between the R peaks.

Respiratory data was processed with similar pipeline of a previous study[4]. First, outliers in respiratory traces were corrected using a moving median filter ("filloutliers.m" function in Matlab). Then, the traces were smoothed with Savitzky-Golay filter using a 1-sec time window ("smoothdata.m" function in Matlab) and z-scored. We then detected respiratory peaks in the smoothed signal using "findpeak.m" function in Matlab (minimum peak distance = 2 seconds and minimum peak prominence = 0.01 seconds), and the respiratory interval was calculated by getting distance between the peaks. The inter-beat-interval and respiratory interval data were downsampled to 25 Hz and converted to heart rate (HR) and respiratory rate (RR), respectively. The HR and RR data were divided into pre-defined numbers of time-bins (5 or 40 bins), and the data for each time-bin was averaged.

fMRI data acquisition. Whole-brain fMRI data were acquired on a 3T Siemens Prisma scanner at the Sungkyunkwan University. The scanning parameters were same with Study 2 except for the number of volumes, which was 2608.

fMRI data analysis. Preprocessing was done similarly as in Studies 1 and 2, but there were also some differences: First, we did not use the automated MRN preprocessing pipeline, but we manually followed the same preprocessing steps using the same tools. Second, we conducted the additional preprocessing including nuisance regression separately for each run because each run was long enough to capture the baseline brain activity. Third, we included the timepoints for the intermittent simple response task as nuisance covariates. Fourth, we removed 22 initial volumes, instead of 20 volumes, to allow enough time for image intensity stabilization.

Study 4: Clinical back pain dataset (subacute and chronic back pain)

Participants. Study 4 data were obtained from the OpenPain Project (OPP) database (http://www.openpain.org/). The OPP project (Principal Investigator: A. Vania Apkarian, Ph.D. at Northwestern University) is supported by the National Institute of Neurological Disorders and Stroke (NINDS) and National Institute of Drug Abuse (NIDA). This dataset consisted of longitudinal fMRI study of clinical back pain patients, including 70 subacute back pain (SBP) patients (age = 43.3 ± 10.6 [mean ± SD], 34 female) and 25 chronic back pain (CBP) patients (age = 44.6 ± 7.9 [mean ± SD], 9 female). All the SBP patients included in this study reported their overall pain level higher than 40 based on visual analogue scale (VAS; 0: no pain, 100: maximum imaginable pain) and their duration of back pain was between 4 to 16 weeks, which had not occurred in the last 12 months before the onset of current pain symptom. Participants with psychiatric, neurological, systemic disorders, or high depression scores (>19 scores based

on Beck's Depression Inventory [BDI])[5] were excluded. For more detailed information about the study, please see ref. [6].

Experimental design. Patients were followed longitudinally with fMRI scanning for 1 to 3 years. There were two types of experimental conditions in this study. The first one was the "spontaneous pain rating" condition, in which the participants provided their moment-by-moment ratings of spontaneous pain by spanning their finger. The second one was the "resting state" condition, in which participants kept their eyes open during the scan without any other tasks. Each run lasted for 10 minutes and 10 seconds, and the overall levels of pain were collected using visual analog scale (VAS) within an hour before the start of the fMRI scan.

fMRI data acquisition. Whole-brain fMRI data were acquired on a 3T Siemens TrioTim scanner. High-resolution T1-weighted structural images were acquired. Functional EPI images were acquired with TR = 2500 ms, TE = 30 ms, 64×64 matrix, 3.4×3.4×3.0 mm^3 voxels, 36 interleaved slices, number of volumes = 244. For more details about the scanning parameters, please see ref. [6].

fMRI data analysis. Resting state fMRI data of the OPP database were preprocessed using the Fusion of Neuroimaging Preprocessing (FuNP) pipeline that integrated AFNI and FSL software[7]. The first 4 volumes were removed to allow for image intensity stabilization. Head motion and slice timing were corrected and intensity normalization of the 4D volumes were applied. The nuisance variables of head motion, white matter, cerebrospinal fluid, cardiac pulsation, and arterial and large vein related contributions were removed using FMRIB's ICA-based X-noiseifier (ICA-FIX) software. The fMRI data were registered onto the preprocessed T1-weighted data and subsequently onto the Montreal Neurological Institute (MNI) 3 mm^3 standard space. The low-pass filter with 0.1 Hz and spatial smoothing with full width at half maximum (FWHM) of 4 mm were applied.

Study 5: Clinical back pain dataset (chronic back pain)

Participants. Study 5 was also from the OPP database. This dataset includes fMRI resting state data of CBP patients and healthy aged-matched controls from two independent sites (Japan and UK). The Japan dataset consisted of 24 CBP patients (age = 46.3 ± 11.3 [mean ± SD], 12 female) and 39 healthy control participants (age = 39.1 ± 13.5 [mean ± SD], 14 female), and the UK dataset consisted of 17 CBP patients (age = 44.0 ± 11.4 [mean ± SD], 12 female) and 17 healthy control participants (age = 44.4 ± 11.8 [mean ± SD], 11 female). All the CBP patients included in this study reported duration of back pain more than 12 months. Participants with psychiatric, neurological, or systemic disorders and MRI contraindications were excluded. For more detailed information about the study, please see ref. [8].

Experimental design. Resting state fMRI scan was conducted, where participants kept their eyes open during the scan without any other tasks. Each run lasted for 9 minutes and 45 seconds.

fMRI data acquisition. Whole-brain fMRI data were acquired on a 3T Siemens TrioTim scanner at the CiNet (Osaka, Japan) or Addenbrooke's hospital (Cambridge, UK). High-resolution T1-weighted structural images were acquired. For the Japan dataset, functional EPI images were acquired with TR = 2500 ms, TE = 30 ms, field of view = 212 mm, 64×64 matrix,

3.3×3.3×4.0 mm^3 voxels, 41 ascending-ordered slices, number of volumes = 234; For the UK dataset, functional EPI images were acquired with TR = 2000 ms, TE = 30 ms, field of view = 192 mm, 64×64 matrix, 3.0×3.0×3.8 mm^3 voxels, 32 interleaved slices, number of volumes = 295. For more details, please see ref. [8].

fMRI data analysis. Preprocessing was done with the same pipeline as Study 4.

Study 6: Heat-induced phasic pain dataset

Participants. Thirty-three healthy and right-handed participants were included (age = 27.9 ± 9.0 [mean ± SD], 22 female). Participants were recruited from New York City. The institutional review board of the Columbia University approved the study, and all participants provided written informed consent. Preliminary eligibility of participants was determined through an online questionnaire. Participants with psychiatric, neurological, or systemic disorders and MRI contraindications were excluded. For more detailed description of the inclusion/exclusion criteria, please see ref. [9].

Experimental design. We used thermal stimulation to induce the experimental phasic pain (EPP) in participants. The thermal stimulation was delivered to the volar surface of left inner forearm. Each heat stimulation lasted 12.5 seconds, which consisted of 3 seconds of ramp-up, 7.5 seconds of plateau, and 2 seconds of ramp-down. A total of 6 levels of temperature (44.3°C- 49.3°C, 1°C increment) was used for stimulation. After heat stimulation, participants provided ratings of (i) whether the stimulus was painful or not, and (ii) how much intense the stimulus was using VAS. Intensity ratings for non-painful trials were coded as 0-100, and ratings for painful trials were coded as 100-200. The experiment included 9 runs, of which 7 runs (run 1, 2, 4, 5, 6, 8, and 9) were 'passive experience' runs, and the other 2 runs (run 3, 7) were 'regulation' runs. In this study, we only used the data from the 'passive experience' runs, in which participants were asked to passively experience the painful sensation without any psychological effort to increase or decrease pain. More details about the experimental design are available at ref. [9].

fMRI data acquisition. Whole-brain fMRI data were acquired on a 3T Philips Achieva TX scanner at the University of Colorado Boulder. High-resolution T1-weighted structural images were acquired. Functional EPI images were acquired with TR = 2000 ms, TE = 20 ms, parallel imaging, SENSE factor = 1.5, field of view = 224 mm, 64×64 matrix, 3×3×3 mm^3 voxels, 42 interleaved slices, number of volumes = 213 (for 'passive experience' runs) or 195 (for 'regulation' runs). Stimulus presentation and behavioral data acquisition were controlled using E-Prime software (PST Inc.). For more details about the parameters, please see ref. [9].

fMRI data analysis. Preprocessing was done with the same pipeline as ref. [9]. Structural T1-weighted images were co-registered to the functional image and normalized to MNI. Four initial volumes of fMRI data were removed in order to allow for image intensity stabilization. These functional EPI images were distortion-corrected, motion-corrected (realigned), normalized to MNI using T1 images with the interpolation to 2×2×2 mm^3 voxels, and smoothed with an 8 mm FWHM Gaussian kernel. Then, total 9 runs of preprocessed fMRI images were concatenated, and nuisance covariates described in the 'First-level analysis and robust regression' section of ref. [9] were removed. These nuisance covariates included (i) intercept for each run; (ii) linear drift across time within each run; (iii) 24 head motion parameters (6

movement parameters including x, y, z, roll, pitch, and yaw, their mean- centered squares, their derivatives, and squared derivative)[3]; (iv) indicator vectors for outlier time points identified based on their multivariate distance from the other images in the sample; (v) indicator vectors for the first two images in each run; (vi) signals from white matter and ventricle. Then, high pass filter with 1/180 Hz was applied to the images. For more detailed description about the preprocessing, please see ref. [9].1

Supplementary Data 1: Clinical trial of analgesic treatment dataset (chronic knee pain)

Participants. This dataset was also from the OPP database. It includes fMRI resting state data of 56 chronic knee osteoarthritis patients (age = 57.9 ± 7.0 [mean ± SD], 30 female). All the osteoarthritis patients included in this study reported their overall pain level higher than 4 of 10 based on numerical rating scale (NRS), and their duration of pain was longer than 12 months. These patients met the American College of Rheumatology criteria for osteoarthritis, which was further confirmed by a clinician. For more details about the study, please see ref. [10].

Experimental design. Resting state fMRI data with the clinical trial of analgesic treatment effects were collected. Among the 56 chronic osteoarthritis patients, 37 patients received placebo treatment and the other 19 patients received analgesic drug (duloxetine) treatment. 17 patients of the placebo-treated group were scanned 2 weeks before the treatment. The other 20 patients of the placebo group, and the 19 patients of the duloxetine group were scanned 3 months before the treatment. Each run lasted for 12 minutes and 40 seconds except for one patient who was scanned for 10 minutes and 10 seconds. Analgesic effect was calculated as a percentage of decrease in the overall levels of pain (measured with VAS) from baseline to the end of the treatment period. Patients with analgesic effects higher than 20% were classified as 'responder', and the others were classified as 'non-responder'.

fMRI data acquisition. Whole-brain fMRI data were acquired on a 3T Siemens TrioTim scanner. High-resolution T1-weighted structural images were acquired. Functional EPI images were acquired with TR = 2500 ms, TE = 30 ms, 64×64 matrix, 3.4×3.4×3.0 mm^3 voxels, 40 interleaved slices, number of volumes = 300. For more details, please see ref. [10].

fMRI data analysis. Preprocessing was done with the same pipeline as Study 4.

Supplementary Data 2: Capsaicin-induced tonic pain dataset with gustometer

Participants. Fifty-eight healthy and right-handed participants were included (age = 22.8 ± 2.8 [mean ± SD], 27 female). Participants were recruited from Suwon area in South Korea. As the same with Study 3, the institutional review board of the Sungkyunkwan University approved the study. Preliminary eligibility of participants was determined through an online questionnaire. Participants with psychiatric, neurological, or systemic disorders and MRI contraindications were excluded.

Experimental design. Supplementary Data 2 included five experimental conditions; (i) capsaicin, (ii) bitter taste (quinine), (iii) sweet taste (hot cocoa), (iv) pleasant touch (brush), and (v) control runs. In the current study, we only used data from three conditions, capsaicin, bitter taste, and control. The onsets for the capsaicin and quinine fluid were 1.5 and 7 minutes after the

beginning of the scan, and the stimulus duration for each delivery was 1.5 minutes (Extended Data Fig. 10a). When there was no fluid delivery, water was delivered. Participants were asked to continuously rate their subjective feelings of unpleasantness for the whole run on the general Labeled Magnitude Scale[1] adopted for pleasantness and unpleasantness rating using an MR-compatible trackball device. After each run, we rinse out participants' mouth using water to remove any residual fluid. For control run, we delivered water throughout the run.

fMRI data acquisition. Whole-brain fMRI data were acquired on a 3T Siemens Prisma scanner at the Sungkyunkwan University. The scanning parameters were same with Study 2 except for the number of volumes, which was 1893 (14.5 minute scan for each run).

fMRI data analysis. Preprocessing was done using the same pipeline as Study 3, except for the following differences. First, 18 initial volumes (instead of 22 volumes) were discarded for image intensity stabilization. Second, the functional data was smoothed with a 5 mm FWHM Gaussian kernel (instead of 6 mm FWHM kernel). Third, in addition to 24 head-motion parameters and image-intensity outliers, 5 principal components of white matter and cerebrospinal fluid signals[11] and linear drift were regressed out.

Stimuli and delivery procedure

Capsaicin delivery (Studies 1-3). To induce tonic pain, we applied hot sauce on participants' tongue in Studies 1-3. We used hot sauce (a food ingredient) to induce orofacial tonic pain with a minimal risk. In Studies 1 and 2, we used the Habanero Pepper Sauce from Tabasco®, and in Study 3, we used the Capsaicin Hot Sauce from Jinmifood, Inc. To deliver the hot sauce in the scanner, we dropped a small amount of hot sauce (0.1 ml) onto a filter paper (2 cm * 6.5 cm, rectangle form). We spread the hot sauce in a circle (diameter = 1 cm) on the upper 1/3 of the filter paper. While participants were lying in the scanner, an experimenter handed the filter paper to the participants. The participants carefully placed the capsaicin side of the paper on their tongue (the participants had an opportunity to practice this procedure with a paper without capsaicin inside of the scanner), and closed their mouth. After 30 seconds, we asked the participants to open their mouth and put the paper on the towel on their chest. Participants kept opening their mouth and breathing only through the mouth for one minute to prevent the capsaicin liquid from flowing. In this way, the liquid dries up, and the capsaicin ends up being located in a specific area of the tongue. After one minute, we asked the participants to close their mouth (and keep closing their mouth), while starting the scan. The participants provided their ratings using a MR-compatible trackball device when a rating scale appeared on the screen. The motivation of using this particular delivery procedure includes 1) to reduce the risk of coughing in the scanner, 2) to keep the pain within a tolerable range, while maximizing the pain intensity, and 3) to make the delivery method simple and easy (without additional equipment).

Aversive tastant delivery (Studies 2-3). To induce aversive, but not painful, oral sensation for the specificity testing of the tonic pain model, we used quinine that has a bitter taste in Studies 2-3. A small amount of quinine sulfate (50 mg) was dissolved in distilled water (0.1ml), which was sufficient to induce aversive and bitter taste. The quinine solution was subsequently dropped onto a filter paper, and the rest of this procedure was same as the 'Capsaicin delivery' above.

Aversive odor delivery (Study 3). We used fermented skate as an additional aversive stimulus for specificity testing of the tonic pain model in Study 3. The fermented skate (Hong-eo) is a food in South Korea famous for its bad smell. We chose to use the fermented skate among many other options (including multiple strong-smell cheese, Maroilles Fauquet, Bons Mayennais Lingot, Gorgonzolla Piccante) based on aversiveness ratings in a pilot study (n = 15). We attached a slice of fermented skate, covered with a filter paper, to the interior nasal part of a mask. The mask was designed to cover the nose and the mouth. For the delivery, we first moved the bed out of the scanner and unlocked and lifted the head coil. While the participants breathing through their mouth, we placed the mask to cover participant's nose and mouth. After the head coil was installed again, the participants were re-entered into the scanner. We asked the participants to start breathing through the nose after we started the scanning. We instructed the participants to breathe only through the nose until the end of the scan.

Capsaicin delivery with an MR-compatible gustometer system (Supplementary Data 2). In this study, we used a newly built MR-compatible gustometer system to induce orofacial tonic pain multiple times within a run. The gustometer system delivers fluid through an MR-compatible mouthpiece, the shape of which was adjusted before the scan to minimize discomfort during the scan. The same capsaicin hot sauce (Jinmifood, Inc.) as in Study 3 was used, but this time, we used a 20% concentration of capsaicin fluid by dissolving 20 ml of the capsaicin hot sauce in 80 ml of water because a liquid with high viscosity cannot be used with the gustometer system. The onsets for the capsaicin fluid were 1.5 and 7 minutes after the beginning of the scan, and the stimulus duration for each delivery was 1.5 minutes (Extended Data Fig. 10a). To keep the pain intensity within tolerable limits, the fluid delivery speed was very slow so that the amount of capsaicin fluid per one delivery was less than 1 ml. When there was no capsaicin delivery, water kept flowing out. During the scan, the delivered fluid was removed using a suction pump to prevent participants from swallowing fluid. The computer-controlled 8-channel fluid delivery system, OctaflowⅡ (ALA Scientific Instruments Inc., Westbury, NY), controlled the whole process of fluid delivery.

Quinine delivery with an MR-compatible gustometer system (Supplementary Data 2). To deliver the quinine using the gustometer, we used the 20mM quinine solution. Then, same procedure as above was used for delivering the quinine solution to participants' mouth.

Functional connectivity processing details for each study

Capsaicin tonic pain datasets (Studies 1-3). Both 'overall' and 'time-binned' functional connectivity features were used for Studies 1-3 (n = 109). Temporal binning of Study 1 and 2 data was done based on the time interval between pain ratings. The last time bin was defined as a period between the last pain rating and the end of the scan. The temporal binning of Study 3 data was done by dividing the BOLD timeseries into pre-defined numbers of bins (5 or 10 bins), because the continuous pain rating was used in Study 3.

Clinical pain datasets (Studies 4-5 and Supplementary Data 1). Only the 'overall' functional connectivity features were used for Studies 4-5 and Supplementary Data 1 (n = 248). In study 4, we averaged all the connectivity matrices of multiple longitudinal scans for each task type ('spontaneous pain rating' and "resting state") and for each patient, because our main goal here was to predict the individual differences in overall pain severity.

Heat-induced phasic pain dataset (Study 6). Only the 'time-binned' functional connectivity was used for Study 6 data (n = 33). Given that Study 6 used the event-related experimental design, we calculated functional connectivity patterns for each trial, which served as a time bin. We defined a period of BOLD signal between the 5 seconds before the start of pain stimulation and the 4.5 seconds after the end of pain stimulation, which was total of 22 seconds. Because there were 77 trials in the passive experience condition, the number of time bins we used here was 77 per participant. Functional connectivity was calculated by taking Pearson's correlation or averaging dynamic correlations using DCC for each time bin data.

Capsaicin tonic pain dataset with gustometer (Supplementary Data 2). The 'time-binned' functional connectivity features were used for the Supplementary Data 2 (n = 58). The initial 30 seconds of fMRI data were excluded because they were too noisy. Given that there were two stimulation period within a run (1.5 and 7 minutes from the start of the scan), we used 1 minute data (131 TRs) to construct the first time-bin and 1.8 minute data (239 TRs) to construct the following 7 time-bins, which yielded a total of 8 time-bins. Based on this time-binning scheme, the first bin (0.5-1.5 minutes) corresponded to the baseline period without stimulation, and the second (1.5-3.3 minutes) and fifth (7-8.8 minutes) time-bins corresponded to the first and second stimulation periods. We discarded the last 6 TRs of the scan, to ensure that all the time-bins had same length except for the first time-bin.

SUPPLEMENTARY RESULTS

Behavioral and physiological results (Extended Data Fig. 1)

In Study 1, we scanned 19 participants with and without capsaicin stimulation ("Capsaicin" and "Control" conditions), each for 5.25 minutes. Capsaicin is a major ingredient of hot chili pepper and activates the transient receptor potential vanilloid 1 (TRPV1) channel, which is involved in the transmission of nociceptive signals[12]. We applied capsaicin-rich hot sauce on participants' tongues using a filter paper before the fMRI scan started. During the scans, we asked participants to report pain intensity and unpleasantness ratings on a modified version of the general Labeled Magnitude Scale (gLMS) every 45 seconds[1]. We obtained structural scans between the runs to minimize any remaining sensations, and also provided liquid sucrose until the participant reported no remaining painful sensations on the tongue. The order of capsaicin and control condition was counterbalanced across participants to avoid any confounding order effects.

Behavioral results showed that pain intensity and unpleasantness ratings during the capsaicin condition were significantly higher than those during the control condition for all time-points, for pain intensity ratings, $\hat{\beta} = 0.10 \pm 0.01$ (mean ± standard errors of the mean), $z = 3.84$, $P = 0.0001$; for unpleasantness ratings, $\hat{\beta} = 0.07 \pm 0.01$, $z = 3.61$, $P = 0.0003$, multi-level general linear model with bootstrap tests (10,000 iterations), gender and the order of experimental conditions were modeled as covariates (Extended Data Fig. 1a). This indicates that the capsaicin administration successfully induced tonic pain in participants. We also observed elevated autonomic responses in the capsaicin condition for all time-points, including heart rate and skin conductance response, for heart rate, $\hat{\beta} = 1.62 \pm 1.24$, $z = 1.68$, $P = 0.047$, one-tailed; for skin conductance response, $\hat{\beta} = 0.62 \pm 0.28$, $z = 2.14$, $P = 0.033$, two-tailed (Extended Data Fig. 1b).

Noise analysis (Extended Data Fig. 5)

To examine whether the ToPS performance was influenced by confounding variables such as head motion[13] and physiological noises[14], we conducted multiple noise analyses.

First, we compared the mean framewise displacement (FD)[13], heart rate (HR), and respiratory rate (RR) between the capsaicin versus control conditions with the primary independent test dataset (Study 3; Extended Data Fig. 5a). The mean FD and HR were significantly higher, and the RR was lower in the capsaicin condition compared to the control condition (FD: $t_{47} = 3.60$, $P = 0.0008$; HR: $t_{37} = 5.99$, $P = 6.57 \times 10^{-7}$; RR: $t_{37} = -2.56$, $P = 0.015$, paired *t*-test), suggesting that while participants were experiencing tonic pain, motion and heart rate was increased, but breathing was slowed. Note that we needed to exclude 10 participants' data due to technical issues with physiological data acquisition.

We then used a set of 36 nuisance variables to predict ToPS model scores. As shown in Extended Data Fig. 5b, the model included 24 head movement-related predictors, 10 components from CSF and white matter (WM), heart rate, and respiration rate. This creates a nuisance model optimized to predict ToPS. Low predictive accuracy would constitute evidence that the ToPS model output is independent from nuisance variables. Some relationship between biomarkers designed to track pain intensity and physiological responses is expected, as physiological responses are also known to track pain intensity[15]. However, the critical questions are (1)

whether physiological responses or other nuisance covariates have large enough effects to explain the relationship between ToPS and pain, and (2) whether the ToPS predicts pain controlling for the nuisance covariates. The analyses below demonstrate that the answers to these questions are no and yes, respectively.

As shown in Extended Data Fig. 5b, for the capsaicin condition, the nuisance regressors predicted ToPS model scores with a significant but very small effect: mean prediction-outcome $r = 0.054$, $P = 0.047$, and for the control condition, mean $r = 0.046$, $P = 0.074$, $n = 38$, one-sample t-test. Movement estimates and WM+CSF components alone did not have a significant relationship with the ToPS. This result indicates that physiological variables have a systematic relationship with the ToPS response, but the variance explained is on the order of 0.3%. Thus, the answer to question (1) above is that the nuisance covariates we tested did not have a large enough relationship with the ToPS to explain its relationship with pain.

The second question we tested is whether the ToPS response can explain tonic pain intensity above and beyond (i.e., after controlling for) these physiological variables. Extended Data Fig. 5c shows the results from a multi-level general linear model (GLM) analysis, in which we predicted ratings with the ToPS response, head motion (FD), mean WM and CSF signals, and physiology variables including HR and RR ($n = 38$). The results showed that the ToPS response was a significant predictor of pain ratings above and beyond other variables ($\hat{\beta}_{\mathrm{ToPS}} = 0.06 \pm 0.02$ [mean ± SEM], $z = 5.62$, $P = 1.86 \times 10^{-8}$, multi-level GLM with bootstrap tests, 10,000 iterations). The other nuisance and physiological variables did not significantly predict tonic pain ratings after controlling for the ToPS response ($\hat{\beta}_{\mathrm{FD}} = 0.02 \pm 0.02$, $z = 1.20$, $P = 0.231$; $\hat{\beta}_{\mathrm{WM}} = 0.01 \pm 0.01$, $z = 1.46$, $P = 0.145$; $\hat{\beta}_{\mathrm{CSF}} = -0.01 \pm 0.01$, $z = -1.10$, $P = 0.270$; $\hat{\beta}_{\mathrm{HR}} = 0.03 \pm 0.02$, $z = 1.10$, $P = 0.270$; $\hat{\beta}_{\mathrm{RR}} = -0.01 \pm 0.01$, $z = -1.10$, $P = 0.272$).

Overall, these results suggest that ToPS captures unique variance in tonic pain above and beyond the nuisance and physiology variables.

Robustness to a different preprocessing pipeline

To examine whether the model performance was robust to different preprocessing pipelines, we tested our tonic pain model on the same Study 3 data, but with a different preprocessing pipeline. The new preprocessing pipeline was similar to the Human Connectome Project (HCP) pipeline[16] and was based on AFNI, FSL, and Freesurfer. For structural T1-weighted images, magnetic field bias was corrected and non-brain tissues were removed using Freesurfer. Then, the images were normalized to MNI using FSL. For functional EPI images, initial volumes (22 images) of fMRI images were removed to allow for image intensity stabilization. Then, the images were motion-corrected using AFNI and distortion-corrected using FSL. These EPI images were co-registered to T1-weighted images using boundary-based registration (BBR) technique[17], which used FSL for first registration and Freesurfer for refinement, as in the HCP pipeline. We then removed motion-related signals of the co-registered EPI images using ICA-AROMA[18]. Additional preprocessing modules including (i) removal of nuisance covariates, (ii) linear de-trending, and (iii) low pass filtering with 0.1 Hz were combined and conducted in one step by 3dTproject function in AFNI to avoid introducing unwanted artifacts[19]. We included (i) mean BOLD signals from white matter (WM) and mean BOLD signals from cerebrospinal fluid (CSF)[18], and (ii) time-points when intermittent arousal

maintenance task appeared (total of 20 times) as nuisance covariates. These de-noised EPI images were then normalized to MNI with the interpolation to 2×2×2 mm^3 voxels and smoothed with a 5 mm FWHM Gaussian kernel using FSL.

With these newly preprocessed data, the model performance was slightly decreased, but not substantially: The predictive performances for the within- and between-individual variations in tonic pain ratings were $r = 0.45$ and $r = 0.41$, respectively, $P < 0.0001$ for both predictions, bootstrap test (the original predictive performances were $r = 0.51$ for both cases). In addition, the model discriminated the capsaicin condition from the bitter taste with 67% accuracy, $P = 0.029$ and from the aversive odor with 85% accuracy, $P < 0.0001$ (the original accuracy = 85% for both contrasts, $P < 0.0001$).

The Tonic Pain Signature performed better in predicting clinical back pain than models derived from clinical back pain datasets themselves (Extended Data Fig. 6)

We further compared the predictive performances of the ToPS with predictive models derived from the clinical pain datasets themselves, i.e., clinical SBP and CBP models (Extended Data Fig. 6). We used the same modeling options as in the ToPS (i.e., Brainnetome parcellations, DCC, and PCR), and the optimal hyper-parameters (e.g., the number of principal components) were determined based on the cross-validation performance. The clinical SBP model was trained using half of the SBP dataset in the spontaneous pain rating condition (training set, $n = 35$) and tested on the remaining half (hold-out set, $n = 35$). The clinical CBP model was trained using resting state scans from the full CBP dataset, due to its smaller sample size ($n = 17$, after excluding fMRI images that had insufficient brain coverage), and tested with cross-validation. For more details about the modeling, please see the Supplementary Information.

The SBP model showed reasonable predictive performance in the training set ($n = 35$, $r = 0.59$, $P = 0.0002$, leave-one-subject-out [LOSO] cross-validation). However, when the model was tested onto the hold-out set ($n = 35$), its performance became worse ($r = 0.36$, $P = 0.032$) than the predictive performance of the ToPS. Even if the ToPS was tested onto the same hold-out data for a direct comparison, the ToPS performance was still better ($r = 0.48$, $P = 0.003$) than the SBP model, though the difference was not statistically significant, $z = 0.78$, $P = 0.433$. The CBP model failed to show significant predictive performance in the training dataset ($n = 17$, $r = -0.28$, $P = 0.269$, LOSO cross-validation). When we tested the ToPS on the same training data for a direct comparison, the ToPS performance was also better ($r = 0.51$, $P = 0.037$) than the CBP model, and the difference was statistically significant, $z = 2.25$, $P = 0.024$.

Thus, the tonic pain model outperforms the models trained on their own clinical pain datasets. These counterintuitive results may be due to the heterogeneous and noisy nature of clinical pain data, highlighting the importance of developing good experimental models, in more tightly controlled conditions, that can serve as a proxy for clinical conditions.

Ventral striatum seed-based connectivity analysis

In Fig. 5, we found that the functional connectivity between the dorsomedial prefrontal cortex (dmPFC) and ventral striatum plays an important role in tracking tonic pain intensity. Given that the ventral striatum and its connections to different medial prefrontal regions have

been implicated in multiple aspects of pain, including pain relief-related learning[20-22], cognitive self-regulation[9], transition to chronic pain[23], reduced motivation in chronic pain[24] and so on, we further examined their functional roles in tonic pain using seed-based connectivity analysis with Study 1 data (n = 19). More specifically, using the bilateral ventral striatum ROIs from the ToPS model (MNI center coordinates: left [-18, 4, -8] and right [14, 8, -8]) as a seed, we calculated the whole-brain seed-based functional connectivity map for each time bin. Similar to our main analysis, 7 time-bins were used.

We first conducted the univariate regression analysis, in which we regressed the time-binned connectivity maps (voxel-wise) on tonic pain intensity ratings across capsaicin and control conditions, and performed the second-level t-tests on the beta maps, treating participants as a random effect. As shown in Extended Data Fig. 8b, the univariate analysis results thresholded at $q < 0.05$ (false discovery rate-corrected) showed that the lateral and dorsal medial prefrontal cortices—including dorsolateral, ventrolateral, and dorsomedial prefrontal cortex, the dorsal part of the ventromedial prefrontal cortex, supplementary motor area—and anterior insula have positive coefficients, meaning that connectivity with the VS is positively associated with tonic pain. In addition, some basal ganglia regions, such as putamen and caudate head, right dorsal posterior insula, and periaqueductal gray (PAG) showed positive coefficients. Conversely, posterior and ventromedial cortical areas showed negative coefficients, including medial and mid-lateral orbitofrontal cortex, S2, and multiple temporal cortex regions. Some brainstem regions, such as pons, also showed negative coefficients.

We next examined whether this pattern was also observed in the multivariate analysis (Extended Data Fig. 8c), in which we used the principal component regression (PCR) with a reduced number of principal components to predict pain intensity ratings based on ventral striatum seed-based connectivity across capsaicin and control condition. The number of principal components was selected based on cross-validated within-individual predictive performance (#PC = 45; mean prediction-outcome r = 0.25, P = 0.002, bootstrap test). To identify important brain regions, we conducted the bootstrap test for the PCR with 10,000 iterations. Note that this prediction performance may be biased given that the hyperparameter, i.e., the number of principal components, was optimized based on this dataset. However, because our goal here is the interpretation of the predictive weights, not the model development, we are presenting this model's predictive weights. The multivariate analysis results showed significant weights within only a few brain regions (Extended Data Fig. 8c; none survived FDR correction), suggesting that in order to generate a robust model, we may need whole-brain connectivity data. When we examined the unthresholded univariate and multivariate regression weights within the dorsomedial and ventromedial prefrontal cortices (dmPFC and vmPFC; Extended Data Fig. 8c), the dmPFC had a higher proportion of positive weights than the vmPFC in both univariate and multivariate maps. This result corroborates the findings shown in Fig. 5 of the main manuscript, and is also consistent with a recent study which identified the dmPFC as a pro-pain brain region that makes extra-nociceptive contributions to pain perception[25].

These results suggest that there is the dorsal-to-ventral gradient of the pain predictive weights in the ventral striatal-medial prefrontal connectivity.

Predicting drug- and placebo-induced analgesic responses to the Tonic Pain Signature (Supplementary Fig. 8)

We tested whether the ToPS could prospectively predict the analgesic responses to placebo pills and a verum drug (Duloxetine) in chronic knee pain patients using the baseline resting state fMRI data[10] (Supplementary Data 1, $n = 56$; also available at openpain.org). As shown in Supplementary Fig. 8, the ToPS was predictive of the future analgesic responses to placebo in the placebo-responder group ($n = 18$, patients who experienced placebo-induced analgesic response > 20% from the baseline level of pain)—the correlation between the predicted and actual outcomes was $r = 0.67$, $P = 0.003$, one-sample t-test. Conversely, the ToPS was not predictive of the analgesic responses to duloxetine in the drug-responder group ($n = 8$, patients who experienced drug-induced analgesic response > 20% from the baseline level of pain) with the prediction-outcome correlation of $r = -0.33$, $P = 0.424$.

To examine the importance of features for the prediction, we used a "virtual lesion analysis," which demonstrated its advantage for examining feature importance of fMRI pattern-based model in a previous study[26]. In the current study, virtual lesions were made by removing a chunk of connections between each functional network and whole brain regions. We tested the predictive performances of the reduced models, and calculated their decreases. A large decrease in the model performance indicates the virtually lesioned features (i.e., functional network) are important for the model prediction. We also tested statistical significance of the feature importance using bootstrap tests. Results (Supplementary Fig. 8) showed that the fronto-parietal network of the model was important for the prediction of analgesic response magnitude in the placebo-responder group, $z = 2.74$, $P = 0.006$, bootstrap test, suggesting that the model component related to top-down cognitive regulation may play an important role in predicting future placebo analgesia. The ToPS was not predictive of changes in knee pain ratings for the non-responder groups ($n = 30$), $r = -0.13$, $P = 0.484$.

Univariate analysis (Supplementary Fig. 9)

To identify the brain activation patterns correlated with tonic pain ratings, we conducted general linear model analysis using the pain intensity and unpleasantness ratings as independent variables (Study 1, $n = 19$). For the brain data, we divided the timeseries of voxel-wise fMRI activity into pre-defined time-bins (as described in Functional connectivity processing), averaged the fMRI activation data within each bin, and concatenated the data across the capsaicin and control runs. For each individual, we regressed binned voxel-wise fMRI activity (Y) on pain intensity or unpleasantness ratings (X) and performed the second-level one sample t-tests on the beta maps, treating participants as a random effect. The statistical maps showed weak (voxel-wise $P < .01$, uncorrected), but sensible results in some expected brain regions, including bilateral sensory-motor regions for mouth/tongue/face, mid- to dorsal-posterior insula (dpINS), and secondary somatosensory/gustatory cortex[27-29].

Caveats of the current study

First, the time-course of pain experience induced by our experimental stimuli is distinct from the typical time-course of clinical pain, which usually fluctuates over time[30]. Whether the ToPS can generalize to pain with different time-courses becomes an important consideration. Though this is certainly an empirical question, the major goal of this study was to model the rise and fall of pain within individuals, and therefore even though the exact time-course of pain

differs across individuals or pain conditions, we would expect our component process marker to track its rise and fall. To test this prediction, we additionally tested the ToPS on a recently collected, and previously unpublished, tonic pain dataset (Supplementary Data 2, $n = 58$) that included a different time-course of oral capsaicin pain. In this study, we used a new tastant delivery system (gustometer) to safely deliver capsaicin into participants' mouths through an MR-compatible mouthpiece (for more details, see Supplementary Information). This new delivery system allows us to induce two or more distinct periods of capsaicin pain within one run, providing an opportunity to test the generalizability of our model to tonic pain experiences with a different time-course. As shown in Extended Data Fig. 10, the ToPS significantly predicted pain, mean $r = 0.33$, $P = 3.32 \times 10^{-9}$ in a bootstrap test. Importantly, the ToPS response also showed two peaks at the time of capsaicin delivery and maximal pain, providing evidence for the generalizability of ToPS to pain with varying temporal patterns.

Second, in the current study, we used only one stimulus modality, oral capsaicin, but there are multiple different types of tonic pain stimuli, including cold water[31], hypertonic saline[32], incision[33], and heat pain under hyperalgesic conditions (e.g., capsaicin)[34], each of which has a unique perceptual quality and spatiotemporal characteristics. Future studies must examine whether our tonic pain model can generalize to other tonic pain modalities, and whether other types of tonic stimuli have reliably distinct brain correlates. A single model might not be sufficient to explain all aspects of pain[25,35] – though we provide evidence here that the ToPS generalizes to chronic back pain, so it may not vary extensively across modalities. To further test this, we should collaboratively achieve a deeper and integrative understanding of the brain representations of sustained pain by (1) developing multiple tonic pain models using diverse tonic pain modalities, (2) examining model generalization across types of pain, and (3) investigating shared and distinct representations among multiple pain-predictive models. To promote this process, we have made the model and code freely available through https://github.com/cocoanlab/tops, so that it can be applied and tested on new datasets.

Third, the ToPS showed a small, but significant, prediction performance for the bitter taste condition in Study 3 (prediction-outcome $r = 0.21$, $P = 0.014$, Supplementary Fig. 3 and Extended Data Fig. 4a). This could be because of some shared features across the capsaicin and bitter taste conditions. For example, unlike aversive odors, both capsaicin and bitter taste conditions use liquid tastants as stimuli. The combination of aversiveness and stimulus modality may provide some shared signal in their fMRI data. However, when we tested the ToPS on the new additional dataset that also included a bitter taste condition (Supplementary Data 2, $n = 58$), the ToPS response showed non-significant correlation with the bitter taste ratings, $r = 0.02$, $P = 0.710$, while it still showed significant prediction for the tonic pain ratings (as described above; Extended Data Fig. 10). This new set of results provides additional evidence for the specificity of the ToPS.

Fourth, diagnosing pain or replacing self-report is not the main goal of the ToPS. Rather, the model is designed to assess one component process, sustained pain experience, which is among multiple components that constitute clinical pain syndromes. If clinical pain is modeled as a whole, it becomes unclear what is being measured because of the multidimensional nature of clinical pain. The advantage of our "component process approach"[35] is that we can specify which components are being measured by experimentally manipulating and computationally modeling a specific component of interest. This characteristic of our signature enables multiple potential clinical applications, which are described in more detail in 'Potential clinical scenarios and

challenges' section. Briefly, the potential clinical uses of the ToPS include the evaluation of analgesic drug effects, tracking the progress of chronic pain, multidimensional assessment and subtyping of patients, providing therapeutic targets, and measuring pain in non-communicative individuals (e.g., infants, patients under anesthesia, or individuals with impaired communication due to mental disabilities or dementia), among others. These map onto several categories identified as use cases for biomarkers by the U.S. Food and Drug Administration[36] and other international agencies[37], and as high-priority targets in recent U.S. National Institutes of Health initiatives[37]. However, clinical use of the ToPS would be premature until the model is fully validated, characterized, and understood in the target clinical populations to which it will be applied[38].

Potential clinical scenarios and challenges

One of the major strengths of our newly developed brain-based biomarker, tonic pain signature (ToPS), is that it can assess one specific component process, sustained pain, which is among many components that constitute clinical pain. If clinical pain is modeled as a whole, it becomes unclear what is being measured because of the multidimensional nature of clinical pain. The advantage of our "component process approach"[35] is that we can specify which component are being measured by experimentally manipulating and computationally modeling a specific component of interest. This characteristic of our signature enables multiple clinical scenarios as described below.

1) *Evaluating the effects of new treatments:* One of the biggest obstacles in developing analgesic drugs is the complexity of processes and contexts that influence pain self-reports. Current practice relies almost exclusively on self-report to evaluate analgesic drug effects, but self-report is influenced by many different pain-related processes, such as sensory, affective, evaluative, social, and motivational factors. It also changes as a function of reporting context, i.e., who is being reported to and in what setting. It is influenced by culture[39], gender roles[40], and the way people interpret scale anchors[41]. One of the major roles of biological measures is as indicators of pathophysiology—biological processes linked to disease. Drugs can then be assessed for effects on these disease-linked biological processes. In early-stage (e.g., Phase 2a) clinical trials, such biomarkers can be used to make early stop/go decisions[42-44]. If drugs do not affect their intended mechanistic targets, they could be considered poor candidates for costly Phase 3 clinical trials in which failures can be financially catastrophic. With the ToPS, we can assess whether a drug has a measurable pharmacodynamic effect on brain systems selectively linked to sustained pain, a primary symptom of disease. This could provide useful information for the evaluation of drug effects and potentially speed up the drug development cycle in clinical trials. The same logic applies to other treatments currently being developed to target chronic pain, including Transcranial Magnetic Stimulation (TMS)[45,46], Transcranial Direct Current Stimulation (tDCS)[47], and neurofeedback[48,49].

2) *Monitoring and decision-making for treatment:* A related but distinct use case of biomarkers, called "monitoring biomarker" (by FDA BEST criteria[36]). Monitoring biomarkers can help track disease progression over time, providing information on when additional intervention is needed. Examples include prostate-specific antigen [PSA] for prostate cancer[50] or carbohydrate antigen 19-9 [CA19-9] for pancreatic cancer[51]. Given that the ToPS was trained to track within-individual changes in pain ratings, it has a potential to serve as a monitoring biomarker for clinical pain conditions. The possible use cases include helping make decisions on (1) the prescription of high-risk analgesic drugs (e.g., opioid), (2) appropriate dosage of analgesic drugs, and (3) planning invasive treatments (e.g., nerve blocks, radiofrequency ablation, dorsal root entry zone lesioning, etc.), all of which are highly needed in current clinical practice. The ToPS could eventually help indicate when such interventions are needed and monitor outcomes to help ascertain the effects of treatments.

3) *Multidimensional profiling and subtyping of patients:* The real advantage of any biomarker is that it can measure biological processes that we could not measure otherwise. If we can develop brain-based biomarkers for the component processes that cannot be easily assessed, these biomarkers will be able to provide useful information to better characterize patients' pain symptoms. Particularly, given that clinical pain conditions are complex and multidimensional by nature, we envision the ToPS as part of a multi-modal assessment of pain. These multiple biomarkers can then serve as intermediate features that are altered in various combinations in different clinical pain conditions, allowing us to do the multidimensional assessment of pain based on the interpretable neurobiological components. Further, these component pain biomarkers could be used for brain-based subtyping (or 'biotypes') of patients with differential disease courses or treatment responses. Again, note that this implementation requires further research, in which we need to characterize the ToPS as part of a multi-modal pain assessment strategy and develop multiple brain-based biomarkers targeting different component processes. This is a longer-term strategy, but could ultimately lead to new taxonomies of chronic pain conditions that are more in line with the pathophysiological mechanisms causing pain.

4) *Providing therapeutic targets:* The ToPS model itself could also provide a therapeutic target by identifying hub regions for sustained pain. For example, brain stimulation techniques, including TMS and tDCS, can target key features identified by the ToPS (e.g., dorsal precuneus of frontoparietal network, or paracentral lobule of somatomotor network; Fig. 4). Similarly, real-time fMRI-based neurofeedback could target the ToPS as a whole, or specific sub-networks.

5) *Measuring pain in non-communicating individuals:* Since the current gold-standard measure of pain is self-report, it is difficult to assess pain in patients who cannot verbally report their pain, such as neonates and infants, elderly, patients with locked-in syndrome or loss of consciousness, and those with severe dementia or intellectual

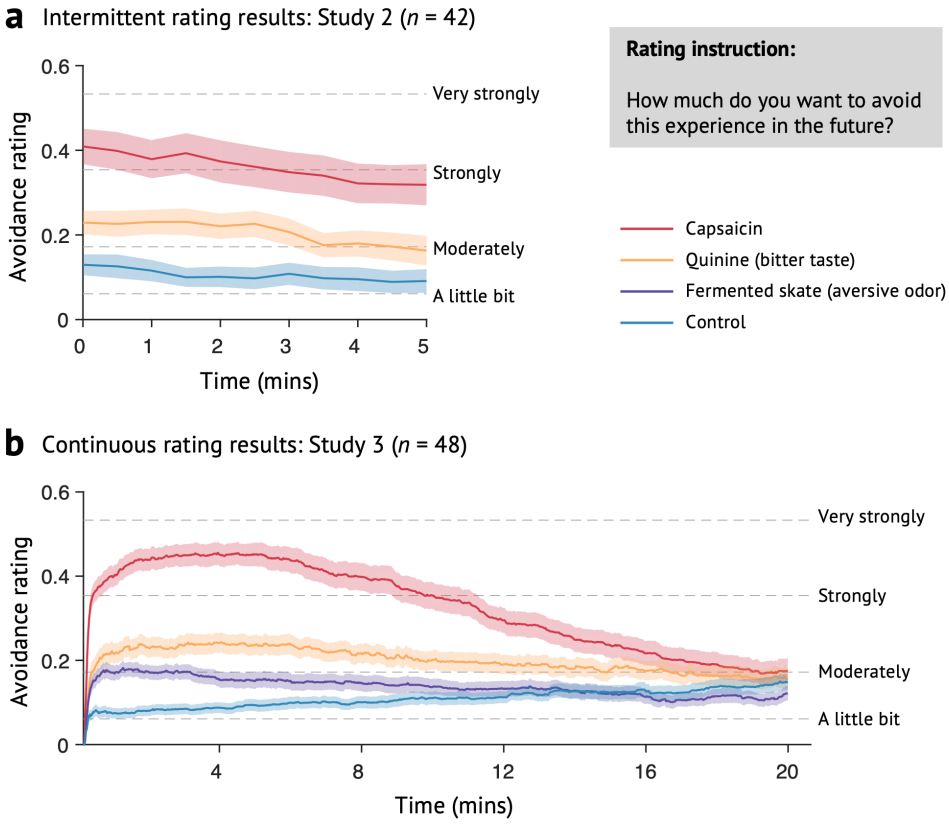
disabilities. The ToPS can serve as a surrogate measure of pain for these populations to help evaluate whether the patients are in pain. Beyond pain detection, the ToPS could provide information on possible causes of pain (i.e., which hubs and subsystems are most affected in an individual), and could also be used to identify optimal dosage of analgesic drugs. Currently, pain pharmacodynamics is very limited in these populations, as it is difficult to define a dose-response curve in the absence of pain[52-54].

There will be several challenges that have to be addressed to realize the proposed clinical implementations. First, there could be some infrastructure issues related to technical and computational requirements. To help address these, we provided a user-friendly tutorial of testing our signature on any preprocessed datasets (https://github.com/cocoanlab/tops). In the future, we should be able to develop open-source software that can take raw data directly from the scanner to provide model predictions. This type of software development was almost impossible before, but recent developments in standardized data structures, preprocessing pipelines, and fMRI data sharing platforms, such as Brain Imaging Data Structure (BIDS[55]), fMRIPrep[56], and openneuro.org[57], have greatly facilitated the development of 'turn-key' fMRI solutions. In addition, like many types of technology, with sufficient scientific evidence, the development of turn-key solutions can become appealing to technology companies. This has happened with MRI as a whole, with both hardware and image acquisition sequences developed in research laboratories and adopted and refined by scanner manufacturers, who distributed stable versions on commercial scanners. Software solutions for image viewing and analysis are continually made available to medical personnel in this way as well. We envision this type of implementation for testing of signatures in the future with stable versions of analysis and visualization methods embedded in software.

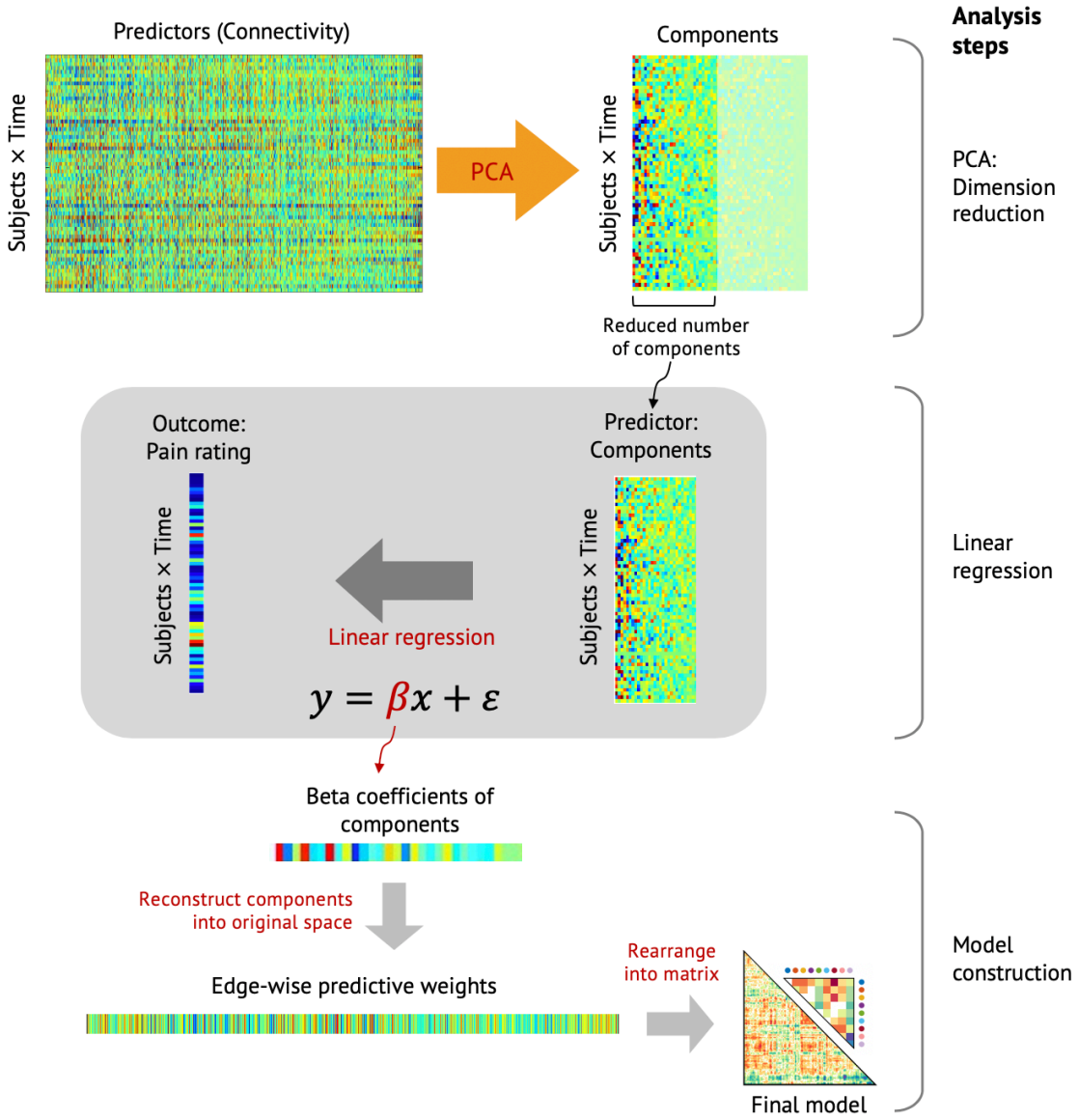
Second, data quality control is an important issue. Since functional connectivity is vulnerable to motion, great care must be taken to minimize in-scanner motion when acquiring fMRI data, especially from those who might have difficulty in following the instructions (e.g., seniors or children). Fortunately, resting-state fMRI is relatively easy to obtain clinically, and there is rapid advancement in standardized quality-control measures, which we integrated into the ToPS pipeline. These take the form of both (a) automated artifact detection and mitigation, and (b) thresholds for automatic rejection of poor-quality data. Given that we demonstrated that the ToPS could predict pain severity of chronic pain patients using resting-state fMRI data, we would recommend using resting-state fMRI in the clinic at this stage.

Third, interpreting fMRI data and analysis results would be one of the biggest challenges to many clinicians who are not familiar with fMRI data or functional networks. Fortunately, our signature is designed to yield a scalar value for an estimate of sustained pain intensity, which is straightforward to interpret and do not require any prior knowledge about fMRI analysis. However, the information the ToPS can offer goes beyond a single numerical value; for example, the ToPS could provide brain-based explanations for individual's pain through analysis methods designed to explain model decisions[58]. For this purpose, we will need additional tool developments that can provide a clinician-friendly visualization of the analyses designed to explain the model and its key features.

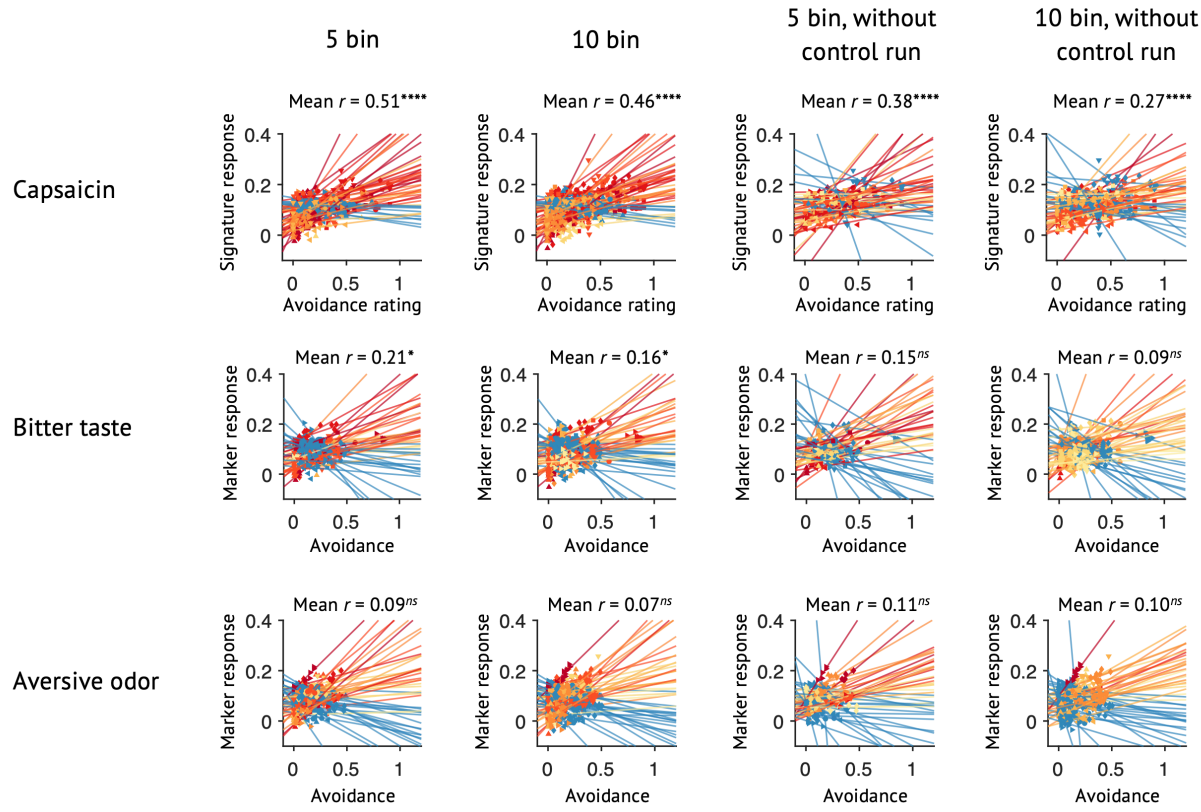
Lastly, one important practical issue will be the cost-benefit of the test. Given that the brain MR Imaging is very expensive, the expected benefits of our signature-based tests should be higher than the cost. However, the cost usually starts high but goes down, sometimes dramatically, as uptake increases (e.g., as has been the case with genetic testing). Through future studies, our signature needs to provide empirical evidence for its clinical utility and prove its benefits.



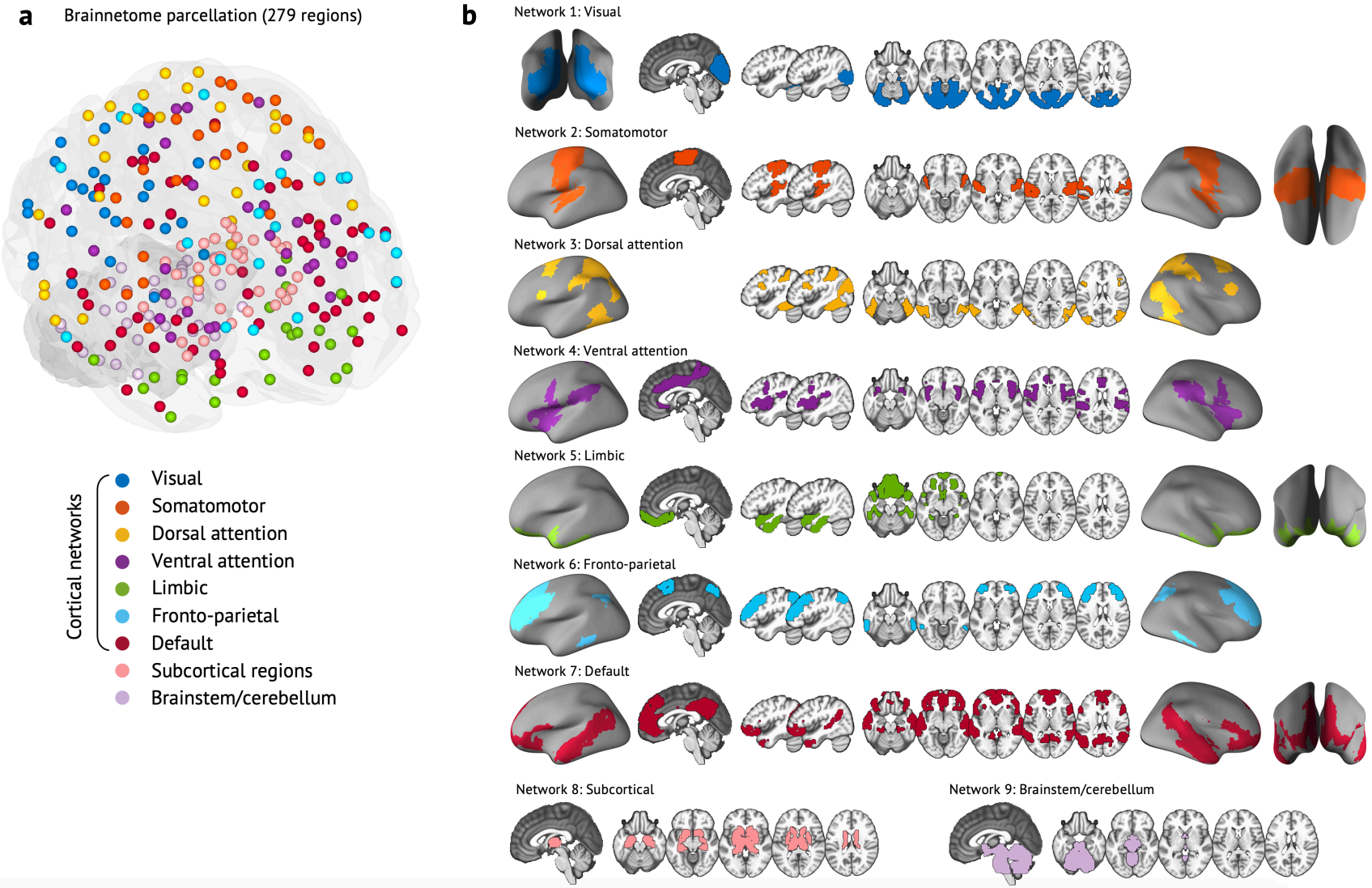
Supplementary Fig. 1. Behavioral data of Studies 2 and 3. To obtain rating scores on a same scale across different stimulus modalities, we used an avoidance rating scale (question: "how much do you want to avoid this experience in the future?") in Studies 2-3. The general Labeled Magnitude Scale (gLMS) was used[1]: The anchors began with "Not at all" [0] to the far left of the scale, and continued to the right in a graded fashion with anchors of "A little bit" [0.061], "Moderately" [0.172], "Strongly" [0.354], and "Very strongly" [0.533], until "Most (I never want to experience this again in my life)" [1] on the far right. Solid lines represent group mean ratings, and shading represents within-subject standard errors of the mean (s.e.m.). We collected avoidance ratings intermittently (every 30 seconds) in Study 2 **(a)** and continuously in Study 3 (**b**).



Supplementary Fig. 2. An overview of the Principal Component Regression (PCR). This figure provides a brief overview of the PCR algorithm that was used in our final model. **(Top)** Dimension reduction using PCA. After we vectorized and concatenated the functional connectivity data, we first reduced its dimensions using Principal Component Analysis (PCA). **(Middle)** Linear regression. Using a reduced number of principal components, we conducted multiple linear regression with pain ratings as a dependent variable. To choose the optimal number of principal components, we did a grid-search method combined with leave-one-participant-out cross-validation. **(Bottom)** Model construction. We projected the beta coefficients obtained from linear regression onto the original space, constructing the final model on the original functional connectivity space.

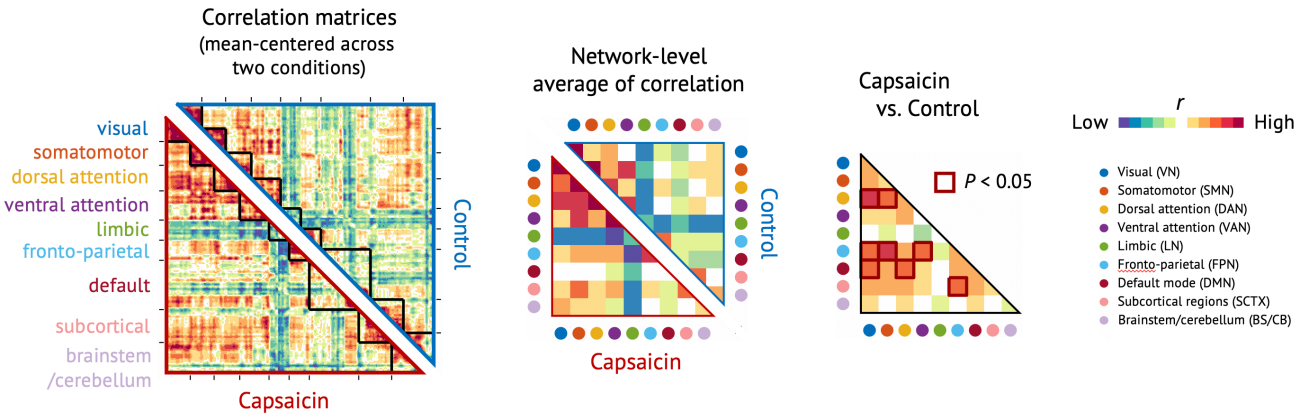


Supplementary Fig. 3. Sensitivity and Specificity tests with or without control run data and with different numbers of time-bins (Study 3). We used the ToPS to predict the avoidance ratings while participants were given capsaicin (first row), bitter taste (second row), or aversive odor (third row). Actual versus predicted ratings (i.e., signature response) are shown in the plots. For the plots of first and second column, each colored line (or symbol) represents an individual participant's data for across the treatment (capsaicin, quinine, or aversive odor) and control runs (red: higher r, yellow: lower r, blue: $r < 0$). For the plots of third and fourth column, each line represents an individual participant's data only for the treatment (capsaicin, quinine, or aversive odor) run. Plots on the first and third columns were based on predictions after averaging within five time bins (5 bins per run), and the those on the second and fourth columns were based on averaging within ten time bins (10 bins per run). The exact P-values for the prediction performance were as follows: For capsaicin (left to right), $P = 3.20 \times 10^{-14}$ (5 bin), 2.73×10^{-14} (10 bin), 6.51×10^{-6} (5 bin, capsaicin only), and 2.82×10^{-5} (10 bin, capsaicin only); For bitter taste, $P = 0.013$ (5 bin), 0.032 (10 bin), 0.053 (5 bin, capsaicin only), and 0.173 (10 bin, capsaicin only); For aversive odor, 0.372 (5 bin), 0.412 (10 bin), 0.156 (5 bin, capsaicin only), and 0.052 (10 bin, capsaicin only), two-tailed, bootstrap tests. ^{ns}non-significant, $^{*}P < .05$, $^{****}P < .0001$.

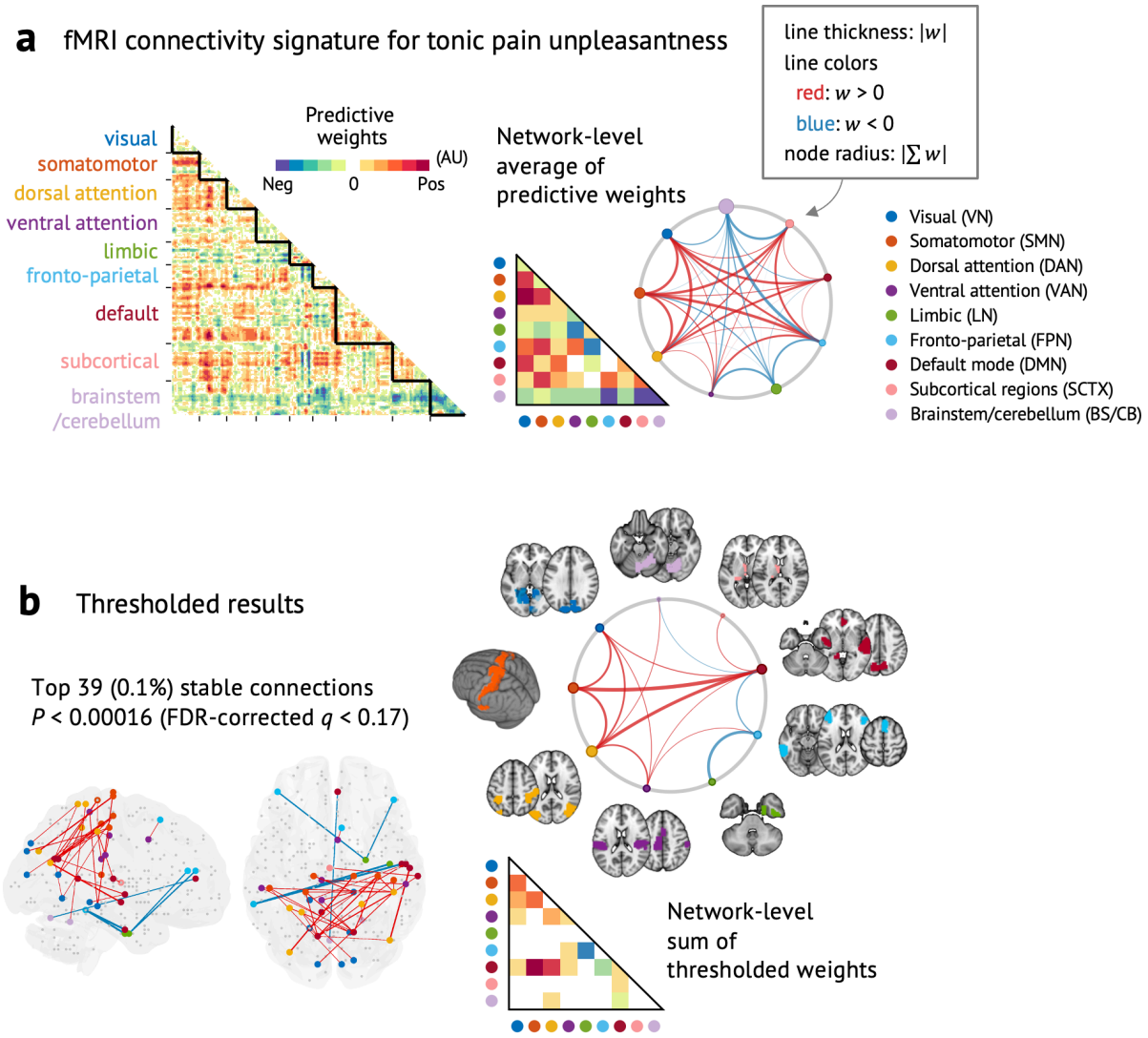


Supplementary Fig. 4. Brainnetome parcellation and functional networks. a, Glass-brain representation of Brainnetome parcellations used in the current study. **b,** To facilitate the functional interpretation of the tonic pain model, we assigned the final brain

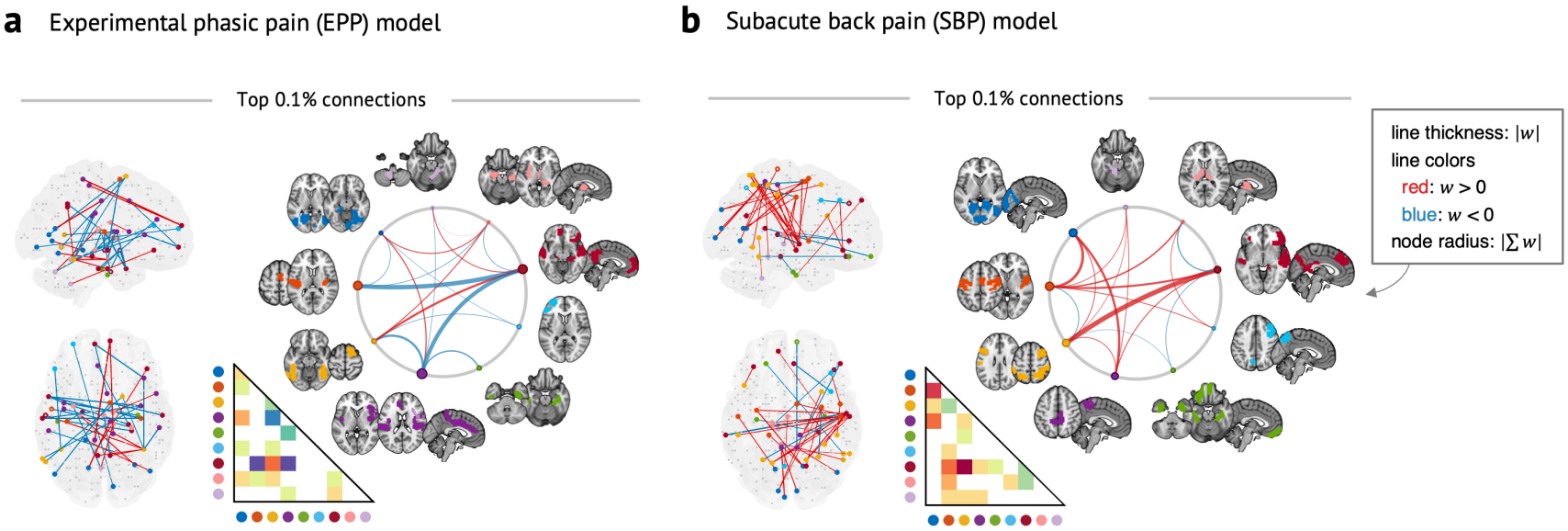
parcellations (279 brain regions) to 9 functional groups, which included 7 cortical functional networks from the Buckner group[59], subcortical regions, and brainstem/cerebellum. The assignment was made based on the number of overlapping voxels between each Brainnetome region and the Buckner's functional atlas combined with manual confirmation and adjustment.



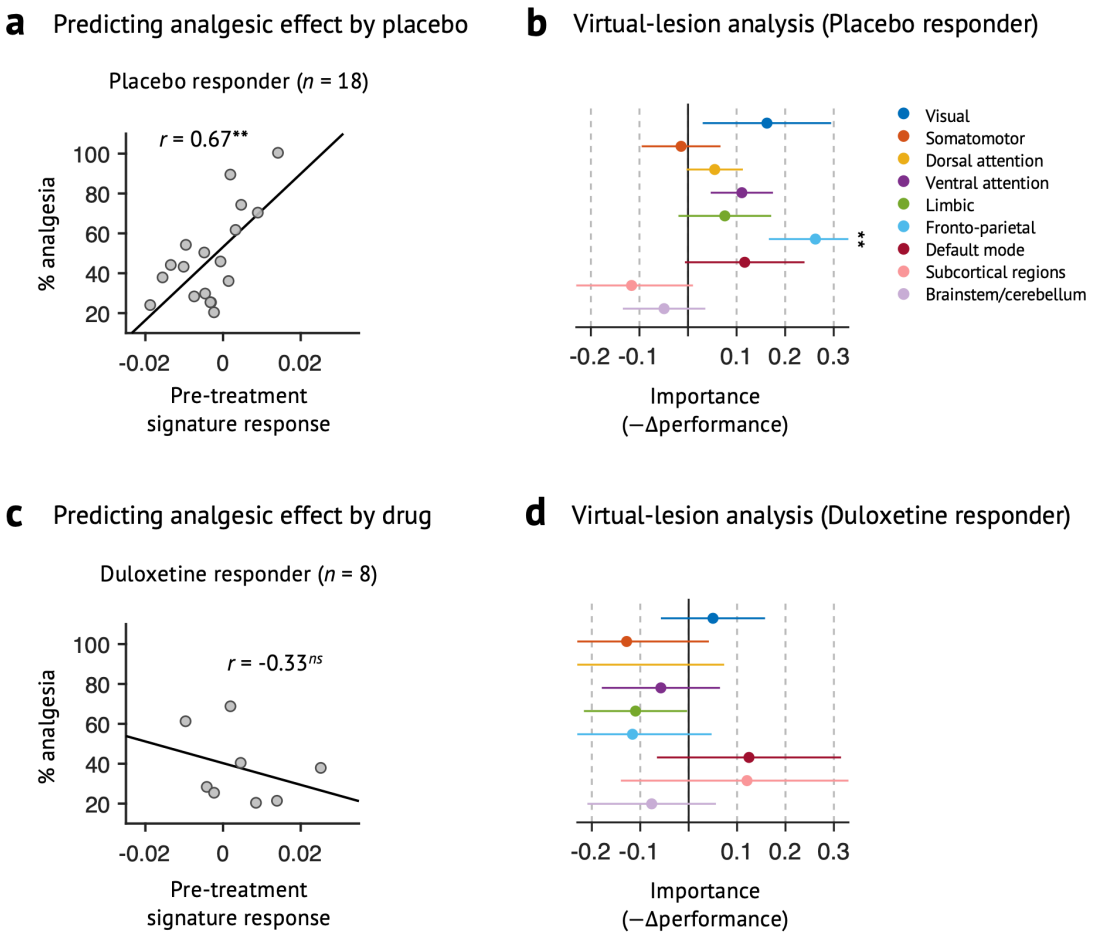
Supplementary Fig. 5. Comparing functional connectivity matrices between the capsaicin and control conditions. The correlation matrices are obtained by averaging dynamic conditional correlations[60] across all TRs within a run and across all participants. To highlight the differences between two connectivity matrices, the correlation matrices were mean-centered across concatenated capsaicin and control conditions. *P*-values were based on a one-sample *t*-test.



Supplementary Fig. 6. Tonic pain "unpleasantness" model. a, Left: The predictive weights of the model. Right: We averaged the model weights for each network-level connection and displayed them with a lower triangular matrix and a circular plot. **b,** We thresholded the map retaining the top 0.1% (39 connections) based on bootstrap tests with 10,000 iterations ($P <$ 0.00016, false discovery rate [FDR] corrected $q < 0.17$, two-tailed). We show the thresholded map using a glass brain and circular plot. All these plots used the same scale for line color, line thickness, and node side to the plots for the pain intensity model in Fig. 4 to help the fair comparisons. Note that comparing the connectivity models between tonic pain intensity versus unpleasantness would also be an interesting and important future direction, but it is beyond the scope of the current manuscript as the study was not designed to independently manipulate intensity and unpleasantness.

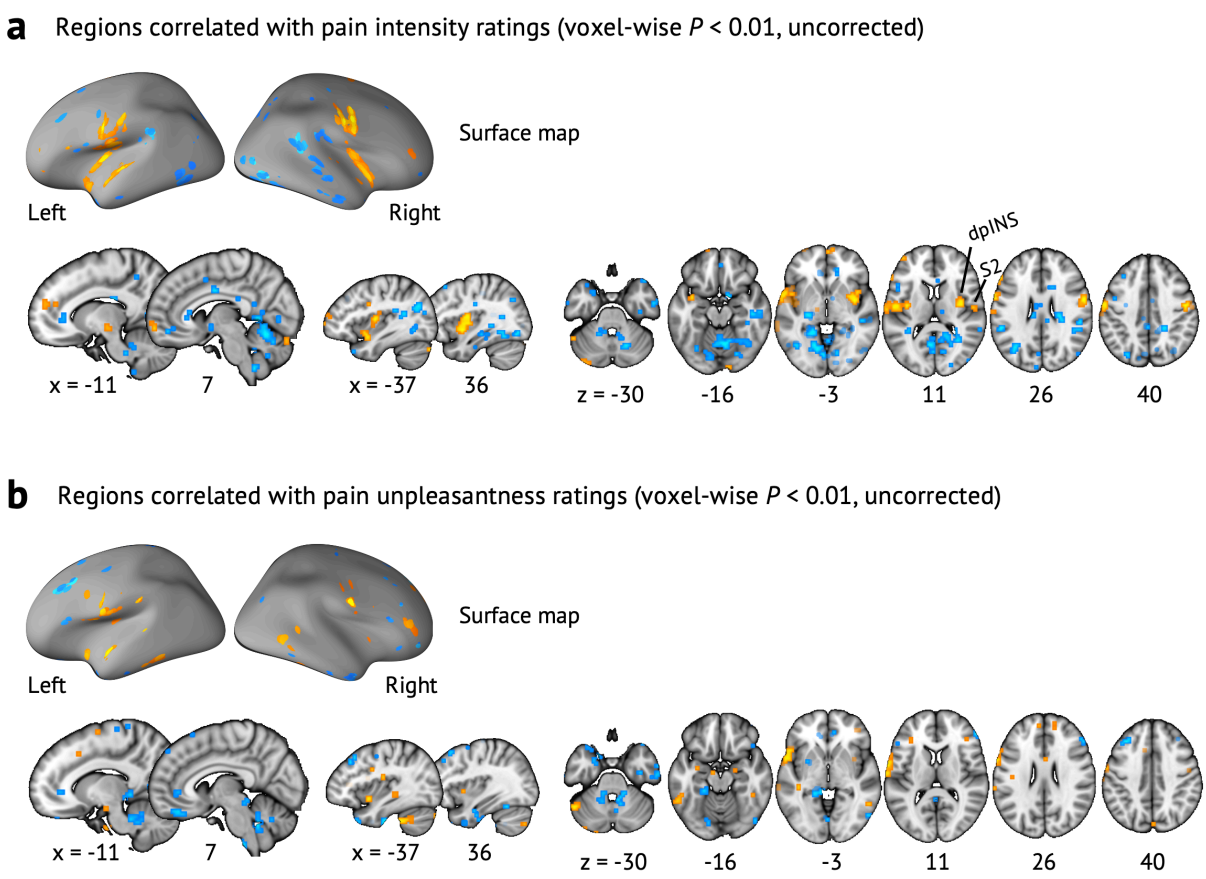


Supplementary Fig. 7. Thresholded experimental phasic pain (EPP) and sub-acute back pain (SBP) models. Thresholded models for **a,** experimental phasic pain induced by heat stimulation and **b,** sub-acute back pain models. The top 39 (0.1%) stable weights were selected using on a bootstrap test with 10,000 iterations. Thresholded connections with a glass brain (left), and network-level sum of connection weights with a circular plot (right) are shown here. We used the same color ranges and line width scales that were obtained from the tonic pain model (shown in Fig. 4).



Supplementary Fig. 8. Prospective prediction of analgesic responses to placebo and drug intervention. We tested the ToPS whether it could predict analgesic responses to placebo pills and a verum drug (Duloxetine) based on pre-treatment resting-state fMRI data in patients with chronic knee osteoarthritis pain (Supplementary Data 1). The data was obtained from a publicly available pain data repository (http://www.openpain.org/) and was previously published[10]. The ToPS response was calculated using the resting state fMRI scans 3 months before randomized-controlled analgesic treatment. **a,** Predicting analgesic response magnitude in 18 patients who responded to placebo treatment (placebo responders). **c,** Predicting analgesic response magnitude in 8 patients who responded to duloxetine (duloxetine responders). The plot shows the relationship between the pre-treatment signature response versus % analgesia (the percentage of pain reduction compared to the baseline). Each dot represents an individual participant, and the line represents the least-square regression line. Note that the signature response at the baseline was highly predictive of analgesic responses to placebo treatment, $r = 0.67$, $P = 0.003$, two-tailed, one sample *t*-test, not drug responses, $r = -0.33$, $P = 0.424$, two-tailed, one sample *t*-test, though the results are expected to be unstable because of the small sample size. **b** and **d,** Virtual-lesion analyses for **b,** placebo treatment and **d,** duloxetine treatment. Each colored dot represents the decreased predictive performance (i.e., -Δ performance) after virtual lesioning of each network, which we defined as importance. The error bars from the center dots represent the standard deviation and the mean of the sampling distribution with bootstrap test. For the placebo-

responder group, the fronto-parietal network showed the significant importance value. $z = 2.74$, $P = 0.006$, two-tailed, bootstrap test. ^{ns}not significant, $^{**}P < 0.01$.



Supplementary Fig. 9. Univariate maps of regions correlated with pain intensity and unpleasantness (Study 1). Brain regions that are correlated with **a,** pain intensity and **b,** pain unpleasantness ratings. Although the current study did not use an experimental design optimized to examine the activation patterns correlated with pain ratings (e.g., event structure, duration, baseline, etc.), we conducted the voxel-wise general linear model (GLM) analysis for a sanity check. For each individual, we regressed voxel-wise fMRI activity averaged over the period between pain ratings (Y) on pain intensity or unpleasantness ratings (X), concatenating across the capsaicin and control runs. Using the beta maps from the first-level GLM analysis, we performed one sample *t*-tests, treating participant as a random effect. dpINS, dorsal posterior insula; S2, secondary somatosensory cortex.

Supplementary Table 1. Predictive performance of the tonic pain signature (pain intensity model) across Studies 1-3

Datasets	Prediction-outcome correlation				Forced-choice classification[c]					
	Within-individual prediction[a]		Between-individual prediction[b]		Capsaicin vs. Control (sensitivity)		Capsaicin vs. Bitter taste (specificity)		Capsaicin vs. Aversive odor (specificity)	
	r	*P*	*r*	*P*	accuracy	*P*	accuracy	*P*	accuracy	*P*
Study 1 (Training, *n* = 19)	0.639	2.99E-12	0.343 (0.285)[d]	0.0350 (0.236)	89%	0.0007				
Study 2 (Validation, *n* = 42)	0.471	3.24E-10	0.446 (0.385)	2.15E-05 (0.0119)	88%	4.43E-07	76%	0.0009		
Study 3 (Independent, *n* = 48)	0.507	3.20E-14	0.515 (0.404)	8.23E-08 (0.0044)	88%	1.01E-07	85%	6.24E-07	85%	6.24E-07

Note. Studies 1-3 included the capsaicin and control conditions that were used for testing model sensitivity. Studies 2 and 3 additionally had the bitter taste and aversive odor conditions that were used for testing model specificity. We used Study 1 as a training set, Study 2 as a validation set, and Study 3 as an independent test set. [a]To evaluate the model performance in predicting within-individual variation in pain ratings, we calculated mean within-individual prediction-outcome correlation. *P*-values were based on a bootstrap test (two-tailed). [b]To evaluate the model performance in predicting between-individual variation in pain ratings, we predicted the average ratings of each condition across individuals. *P*-values were based on one-sample *t*-tests (two-tailed). [c]We also conducted forced choice classification to evaluate the model sensitivity and specificity. *P*-values were based on binomial tests (two-tailed). [d]Numbers in parentheses indicate the between-individual prediction only using the capsaicin condition.

Supplementary Table 2. Prediction performance of the tonic pain unpleasantness model across Studies 1-3

Datasets	Prediction-outcome correlation				Forced-choice discrimination[c]					
	Within-individual prediction[a]		Between-individual prediction[b]		Capsaicin vs. Control (sensitivity)		Capsaicin vs. Bitter taste (specificity)		Capsaicin vs. Aversive odor (specificity)	
	r	*P*	*r*	*P*	accuracy	*P*	accuracy	*P*	accuracy	*P*
Study 1 (Training, *n* = 19)	0.481	6.12E-06	0.264	0.1092	89%	0.0007				
Study 2 (Validation, *n* = 42)	0.507	1.53E-11	0.449	1.88E-05	88%	4.43E-07	79%	0.0003		
Study 3 (Independent, *n* = 48)	0.540	4.82E-19	0.516	7.21E-08	85%	6.24E-07	88%	1.01E-07	90%	1.37E-08

Note. See the Note for Supplementary Table 1.

Supplementary Table 3. Predictive performance of the tonic pain signature on clinical pain data (Studies 4-5, Supplementary Data 1)

Datasets	Clinical populations	Task types	*N*	Performance	
Prediction of overall pain scores[a]				*r*	*P*
Study 4 (Vachon-Presseau et al.)	Subacute back pain	Spontaneous pain rating	53	0.587	3.91E-06
		Resting	53	0.089	0.5283
	Chronic back pain	Spontaneous pain rating	20	0.301	0.1968
		Resting	20	0.558	0.0106
Classification of chronic back pain[b]				accuracy	*P*
Study 5 (Mano et al.)	Chronic back pain (Japan)	Resting	63	73%	0.0003
	Chronic back pain (U.K.)	Resting	34	71%	0.0243
Predicting reduced pain (% analgesia)[c]				*r*	*P*
Supplementary Data 1 (Tetreault et al.)	Placebo responder	Resting	18	0.666	0.0025
	Drug (Duloxetine) responder	Resting	8	-0.330	0.4241
	Non-responder	Resting	30	-0.133	0.4838

Note. We obtained all clinical pain data from the OpenPain Project (OPP) database available at http://www.openpain.org/. [a]Study 4 included longitudinal fMRI scans of patients with sub-acute back pain (SBP) and chronic back pain (CBP) and had multiple different task data[6], including spontaneous pain rating and resting-state tasks. Note that we used the data from patients who had data from both task types (testing $n = 53$ and 20 for SBP and CBP patients, respectively). *P*-values were based on one-sample *t*-tests (two-tailed). [b]Study 5 included resting-state fMRI scans of CBP patients and healthy age-matched controls[8]. The data were collected from two independent sites (Japan and UK). The Japan dataset ($n = 63$) consisted of 24 CBP patients and 39 healthy controls, and the UK dataset ($n = 34$) consisted of 17 CBP patients and 17 healthy controls. *P*-values were based on binomial tests (two-tailed). [c]Supplementary Data 1 included pre-treatment resting-state fMRI data of patients with chronic knee pain (osteoarthritis)[10]. In the study, the patients were randomly assigned to drug (duloxetine) or placebo treatment groups. After the treatment, the patients were classified into treatment responders or non-responders based on a fixed threshold of analgesic response (20% reduction of pain ratings). *P*-values were based on one-sample *t*-tests (two-tailed).

Supplementary Table 4. Top 39 (0.1%) stable connections of the tonic pain signature

Rank	Weights	Regions	MNI coordinates
Positive connections			
#1	0.0003308	Lt. inferior temporal gyrus (BA37, ventrolateral) - Lt. middle occipital gyrus	(-58,-60,-6) - (-34,-87,13)
#2	0.0003259	Lt. middle temporal gyrus (BA37, dorsolateral) - Lt. middle occipital gyrus	(-61,-57,7) - (-34,-87,13)
#3	0.0002891	Lt. precentral gyrus (BA4, trunk) - Rt. precuneus (BA7, medial)	(-16,-21,76) - (3,-63,52)
#4	0.0002799	Lt. precuneus (BA7, medial) - Lt. postcentral gyrus (BA1/2/3, trunk)	(-7,-63,52) - (-22,-33,70)
#5	0.0002773	Rt. inferior parietal lobule (BA40, rostrodorsal) - Lt. parietooccipital sulcus (dorsomedial)	(45,-33,46) - (-13,-66,25)
#6	0.0002771	Rt. precentral gyrus (BA4, upper limb) - Lt. inferior parietal lobule (BA39, rostrodorsal)	(33,-21,58) - (-40,-60,46)
#7	0.0002635	Lt. precentral gyrus (BA4, trunk) - Rt. medial superior occipital gyrus	(-16,-21,76) - (15,-84,37)
#8	0.0002634	Lt. paracentral lobule (BA4, lower limb) - Rt. precuneus (BA7, medial)	(-7,-21,61) - (3,-63,52)
#9	0.0002541	Lt. precentral gyrus (BA4, trunk) - Lt. precuneus (BA7, medial)	(-16,-21,76) - (-7,-63,52)
#10	0.0002474	Rt. precuneus (BA7, medial) - Lt. postcentral gyrus (BA1/2/3, trunk)	(3,-63,52) - (-22,-33,70)
#11	0.0002400	Lt. superior temporal gyrus (BA22, rostral) - Lt. middle temporal gyrus (BA37, dorsolateral)	(-58,-3,-9) - (-61,-57,7)
#12	0.0002322	Lt. precentral gyrus (BA4, trunk) - Lt. superior parietal lobule (BA7, rostral)	(-16,-21,76) - (-19,-60,64)
#13	0.0002285	Lt. paracentral lobule (BA4, lower limb) - Lt. inferior temporal gyrus (BA37, ventrolateral)	(-7,-21,61) - (-58,-60,-6)
#14	0.0002265	Rt. posterior superior temporal sulcus (caudal) - Rt. inferior parietal lobule (BA39, rostrodorsal)	(54,-39,13) - (36,-63,43)
#15	0.0002182	Rt. paracentral lobule (BA4, lower limb) - Rt. lingual gyrus (caudal)	(3,-21,61) - (9,-84,-6)
#16	0.0002085	Rt. inferior parietal lobule (BA40, rostroventral) - Lt. precuneus (BA7, medial)	(54,-27,28) - (-7,-63,52)
#17	0.0001976	Rt. superior temporal gyrus (BA41/42) - Lt. caudate (dorsal)	(51,-24,13) - (-16,4,16)
#18	0.0001834	Rt. posterior superior temporal sulcus (caudal) - Hypothalamus	(54,-39,13) - (-1,1,-9)
#19	0.0001829	Lt. precentral gyrus (BA4, trunk) - Rt. superior parietal lobule (BA7, caudal)	(-16,-21,76) - (15,-69,55)

(Supplementary Table 4 continues on next page)

Rank	Weights	Regions	MNI coordinates
			(Continued from previous page)
#20	0.0001754	Lt. inferior parietal lobule (BA39, rostrodorsal) - Lt. medial superior occipital gyrus	(-40,-60,46) - (-13,-87,31)
#21	0.0001744	Lt. superior parietal lobule (BA7, postcentral) - Lt. inferior parietal lobule (BA39, caudal)	(-25,-48,67) - (-34,-78,31)
#22	0.0001726	Rt. paracentral lobule (BA4, lower limb) - Rt. inferior occipital gyrus	(3,-21,61) - (30,-84,-9)
#23	0.0001654	Rt. precuneus (BA5, medial) - Lt. postcentral gyrus (BA1/2/3, trunk)	(6,-45,58) - (-22,-33,70)
#24	0.0001625	Rt. precuneus (BA7, medial) - Lt. V5/MT+	(3,-63,52) - (-49,-72,7)
#25	0.0001584	Rt. superior temporal gyrus (BA41/42) - Lt. thalamus (caudal temporal)	(51,-24,13) - (-13,-21,16)
#26	0.0001518	Lt. middle frontal gyrus (BA6, ventrolateral) - Rt. precentral gyrus (BA4, upper limb)	(-34,4,55) - (33,-21,58)
#27	0.0001333	Rt. superior parietal lobule (BA5, lateral) - Rt. inferior parietal lobule (BA40, rostroventral)	(33,-42,55) - (54,-27,28)
#28	0.0001313	Lt. superior parietal lobule (BA7, intraparietal) - Rt. parietooccipital sulcus (ventromedial)	(-28,-57,55) - (12,-63,13)
#29	0.0001304	Rt. superior temporal gyrus (BA41/42) - Lt. thalamus (rostral temporal)	(51,-24,13) - (-4,-15,7)
Negative connections			
#1	-0.0004139	Rt. superior temporal gyrus (BA38, medial) - Rt. cerebellum (lobule VI)	(30,16,-33) - (21,-54,-24)
#2	-0.0003892	Rt. superior temporal gyrus (BA38, medial) - Rt. lingual gyrus (rostral)	(30,16,-33) - (15,-57,-6)
#3	-0.0003209	Rt. superior temporal gyrus (BA38, medial) - Rt. parahippocampal gyrus (area TL)	(30,16,-33) - (27,-30,-15)
#4	-0.0003077	Lt. superior temporal gyrus (BA22, caudal) - Rt. superior temporal gyrus (BA22, rostral)	(-64,-33,7) - (54,-12,-3)
#5	-0.0003033	Lt. cerebellum (lobule IX) - Brainstem	(-10,-51,-42) - (-1,-24,-27)
#6	-0.0003008	Lt. parahippocampal gyrus (BA35/36, rostral) - Vermis. cerebellum (lobule IX)	(-31,-6,-33) - (-4,-54,-36)
#7	-0.0002895	Rt. superior temporal gyrus (BA38, lateral) - Lt. inferior temporal gyrus (BA20, rostral)	(45,13,-18) - (-46,-3,-39)
#8	-0.0002862	Rt. superior temporal gyrus (BA38, medial) - Rt. parietooccipital sulcus (ventromedial)	(30,16,-33) - (12,-63,13)
#9	-0.0002532	Rt. precentral gyrus (BA4, head and face) - Lt. inferior temporal gyrus (BA20, rostral)	(51,-3,34) - (-46,-3,-39)
#10	-0.0001850	Lt. inferior temporal gyrus (BA20, rostral) - Lt. cingulate gyrus (BA23, caudal)	(-46,-3,-39) - (-10,-24,43)

Note. Top 39 (0.1%) stable connections based on bootstrap tests with 10,000 iterations ($P < 0.000028$, FDR $q < 0.027$, two-tailed).

REFERENCES

1. Bartoshuk, L.M.*, et al.* Valid across-group comparisons with labeled scales: the gLMS versus magnitude matching. *Physiol Behav* **82**, 109-114 (2004).
2. Bockholt, H.J.*, et al.* Mining the mind research network: a novel framework for exploring large scale, heterogeneous translational neuroscience research data sources. *Front Neuroinform* **3**, 36 (2010).
3. Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S. & Turner, R. Movement-related effects in fMRI time-series. *Magn Reson Med* **35**, 346-355 (1996).
4. Power, J.D.*, et al.* Distinctions among real and apparent respiratory motions in human fMRI data. *Neuroimage* **201**, 116041 (2019).
5. Beck, A.T., Ward, C.H., Mendelson, M., Mock, J. & Erbaugh, J. An inventory for measuring depression. *Arch Gen Psychiatry* **4**, 561-571 (1961).
6. Vachon-Presseau, E.*, et al.* Corticolimbic anatomical characteristics predetermine risk for chronic pain. *Brain* **139**, 1958-1970 (2016).
7. Park, B.Y., Byeon, K. & Park, H. FuNP (Fusion of Neuroimaging Preprocessing) Pipelines: A Fully Automated Preprocessing Software for Functional Magnetic Resonance Imaging. *Front Neuroinform* **13**, 5 (2019).
8. Mano, H.*, et al.* Classification and characterisation of brain network changes in chronic back pain: A multicenter study [version 1; referees: 3 approved]. *Wellcome Open Research* **3**(2018).
9. Woo, C.W., Roy, M., Buhle, J.T. & Wager, T.D. Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLoS Biol* **13**, e1002036 (2015).
10. Tetreault, P.*, et al.* Brain Connectivity Predicts Placebo Response across Chronic Pain Clinical Trials. *PLoS Biol* **14**, e1002570 (2016).
11. Muschelli, J.*, et al.* Reduction of motion-related artifacts in resting state fMRI using aCompCor. *Neuroimage* **96**, 22-35 (2014).
12. Caterina, M.J.*, et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816-824 (1997).
13. Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L. & Petersen, S.E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142-2154 (2012).
14. Birn, R.M. The role of physiological noise in resting-state functional connectivity. *Neuroimage* **62**, 864-870 (2012).
15. Matthewson, G.M., Woo, C.W., Reddan, M.C. & Wager, T.D. Cognitive self-regulation influences pain-related physiology. *Pain* **160**, 2338-2349 (2019).
16. Glasser, M.F.*, et al.* The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105-124 (2013).
17. Greve, D.N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63-72 (2009).
18. Pruim, R.H.R.*, et al.* ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* **112**, 267-277 (2015).
19. Lindquist, M.A., Geuter, S., Wager, T.D. & Caffo, B.S. Modular preprocessing pipelines can reintroduce artifacts into fMRI data. *Hum Brain Mapp* **40**, 2358-2376 (2019).
20. Zhang, S.*, et al.* The control of tonic pain by active relief learning. *Elife* **7**(2018).
21. Seymour, B.*, et al.* Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nat Neurosci* **8**, 1234-1240 (2005).

22. Navratilova, E.*, et al.* Pain relief produces negative reinforcement through activation of mesolimbic reward-valuation circuitry. *Proc Natl Acad Sci U S A* **109**, 20709-20713 (2012).
23. Baliki, M.N.*, et al.* Corticostriatal functional connectivity predicts transition to chronic back pain. *Nat Neurosci* **15**, 1117-1119 (2012).
24. Schwartz, N.*, et al.* Chronic pain. Decreased motivation during chronic pain requires long-term depression in the nucleus accumbens. *Science* **345**, 535-542 (2014).
25. Woo, C.W.*, et al.* Quantifying cerebral contributions to pain beyond nociception. *Nat Commun* **8**, 14211 (2017).
26. Chang, L.J., Gianaros, P.J., Manuck, S.B., Krishnan, A. & Wager, T.D. A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. *PLoS Biol* **13**, e1002180 (2015).
27. Apkarian, A.V., Bushnell, M.C., Treede, R.D. & Zubieta, J.K. Human brain mechanisms of pain perception and regulation in health and disease. *Eur J Pain* **9**, 463-484 (2005).
28. Segerdahl, A.R., Themistocleous, A.C., Fido, D., Bennett, D.L. & Tracey, I. A brain-based pain facilitation mechanism contributes to painful diabetic polyneuropathy. *Brain* **141**, 357-364 (2018).
29. Ayoub, L.J., Seminowicz, D.A. & Moayedi, M. A meta-analytic study of experimental and chronic orofacial pain excluding headache disorders. *NeuroImage: Clinical* **20**, 901-912 (2018).
30. Apkarian, A.V., Krauss, B.R., Fredrickson, B.E. & Szeverenyi, N.M. Imaging the pain of low back pain: functional magnetic resonance imaging in combination with monitoring subjective pain perception allows the study of clinical pain states. *Neurosci Lett* **299**, 57-60 (2001).
31. Mitchell, L.A., MacDonald, R.A. & Brodie, E.E. Temperature and the cold pressor test. *J Pain* **5**, 233-237 (2004).
32. Thunberg, J.*, et al.* Brain processing of tonic muscle pain induced by infusion of hypertonic saline. *Eur J Pain* **9**, 185-194 (2005).
33. Kawamata, M.*, et al.* Experimental incision-induced pain in human skin: effects of systemic lidocaine on flare formation and hyperalgesia. *Pain* **100**, 77-89 (2002).
34. Lorenz, J.*, et al.* A unique representation of heat allodynia in the human brain. *Neuron* **35**, 383-393 (2002).
35. Woo, C.W., Chang, L.J., Lindquist, M.A. & Wager, T.D. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci* **20**, 365-377 (2017).
36. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) resource. (2016).
37. Davis, K.D.*, et al.* Discovery and validation of biomarkers to aid the development of safe and effective pain therapeutics: challenges and opportunities. *Nat Rev Neurol* (2020).
38. Davis, K.D.*, et al.* Brain imaging tests for chronic pain: medical, legal and ethical issues and recommendations. *Nat Rev Neurol* **13**, 624-638 (2017).
39. Zatzick, D.F. & Dimsdale, J.E. Cultural variations in response to painful stimuli. *Psychosom Med* **52**, 544-557 (1990).
40. Robinson, M.E.*, et al.* Gender role expectations of pain: relationship to sex differences in pain. *J Pain* **2**, 251-257 (2001).
41. Scott, J. & Huskisson, E.C. Graphic representation of pain. *Pain* **2**, 175-184 (1976).

42. Wise, R.G. & Preston, C. What is the value of human FMRI in CNS drug development? *Drug Discov Today* **15**, 973-980 (2010).
43. Duff, E.P.*, et al.* Learning to identify CNS drug action and efficacy using multistudy fMRI data. *Sci Transl Med* **7**, 274ra216 (2015).
44. Borsook, D., Becerra, L. & Hargreaves, R. A role for fMRI in optimizing CNS drug development. *Nat Rev Drug Discov* **5**, 411-424 (2006).
45. Klein, M.M.*, et al.* Transcranial magnetic stimulation of the brain: guidelines for pain treatment research. *Pain* **156**, 1601-1614 (2015).
46. Short, E.B.*, et al.* Ten sessions of adjunctive left prefrontal rTMS significantly reduces fibromyalgia pain: a randomized, controlled pilot study. *Pain* **152**, 2477-2484 (2011).
47. Fregni, F.*, et al.* A sham-controlled, phase II trial of transcranial direct current stimulation for the treatment of central pain in traumatic spinal cord injury. *Pain* **122**, 197-209 (2006).
48. deCharms, R.C.*, et al.* Control over brain activation and pain learned by using real-time functional MRI. *Proc Natl Acad Sci U S A* **102**, 18626-18631 (2005).
49. Chapin, H., Bagarinao, E. & Mackey, S. Real-time fMRI applied to pain management. *Neurosci Lett* **520**, 174-181 (2012).
50. Roddam, A.W.*, et al.* Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2-10 ng/ml: systematic review and meta-analysis. *Eur Urol* **48**, 386-399; discussion 398-389 (2005).
51. Zhang, Y.*, et al.* Tumor markers CA19-9, CA242 and CEA in the diagnosis of pancreatic cancer: a meta-analysis. *Int J Clin Exp Med* **8**, 11683-11691 (2015).
52. Duff, E.P.*, et al.* Inferring the infant pain experience: a translational fMRI-based signature study. *bioRxiv*, 2020.2004.2001.998864 (2020).
53. de Knegt, N.C.*, et al.* Behavioral pain indicators in people with intellectual disabilities: a systematic review. *J Pain* **14**, 885-896 (2013).
54. Scherder, E.*, et al.* Pain in dementia. *Pain* **145**, 276-278 (2009).
55. Gorgolewski, K.J.*, et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data* **3**, 160044 (2016).
56. Esteban, O.*, et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* **16**, 111-116 (2019).
57. Gorgolewski, K., Esteban, O., Schaefer, G., Wandell, B. & Poldrack, R. OpenNeuro—a free online platform for sharing and analysis of neuroimaging data. *Organization for human brain mapping. Vancouver, Canada* **1677**(2017).
58. Kohoutova, L.*, et al.* Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nat Protoc* **15**, 1399-1435 (2020).
59. Yeo, B.T.*, et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* **106**, 1125-1165 (2011).
60. Lindquist, M.A., Xu, Y., Nebel, M.B. & Caffo, B.S. Evaluating dynamic bivariate correlations in resting-state fMRI: a comparison study and a new approach. *Neuroimage* **101**, 531-546 (2014).