This document provides additional detail on the procedural and statistical methods and the results, including activation tables and figures that could not be included in the main paper due to space limitations. The description below is intended to supplement the description provided in the paper rather than duplicating material already reported.

Supporting Methods

Participants

Volunteers were 15 healthy right-handed males between 20 and 30 years old. All volunteers were non-smokers who had no personal history of medical, psychiatric illness, substance abuse or dependence, and no family history of inheritable illnesses. Volunteers were not taking psychotropic medications or hormone treatments and did not exercise in excess of 1 hour three times a week. Volunteers were instructed not to drink alcohol for at least 24 hours, nor to exercise or eat for at least 3 hours before the study. Written informed consent was obtained in all cases, and all procedures used were approved by the University of Michigan Institutional Review Board and the Radiation Drug Research Committee. To balance power to detect placebo effects in the group and look for opioid-reported placebo analgesia correlations, 8 participants with confirmed reported placebo analgesia effects were selected from a larger sample reported elsewhere (1), and the remaining 7 participants were recruited de novo. We did not find any effects of returning group status on reported placebo analgesia (see Results below) and thus we collapsed across returning and new participants in other analyses.

Experimental procedures

Stimulation sites. Prior to calibration, testing for warmth-insensitive fields was performed on possible skin sites using a ~43 degree stimulus held against the skin by the experimenter, and skin patches for which participants reported noticeably reduced sensation were excluded. Five eligible 2 cm x 2 cm square regions were marked prior to the study on the left volar forearm. A central site was designated for calibration. Two sites proximal to the body from the calibration site and two sites distal to the calibration site were designated for placebo and control cream application, with the location of the placebo cream counterbalanced across participants. Of the two sites within each of the proximal and distal regions, one was designated for manipulation trials, and the other was designated for test trials, so that each block of trials was administered on a previously unstimulated skin site. This design precludes the possibility that stimuation history during the manipulation phase affects nociception on test blocks.

Calibration. Thermal stimuli were administered in ascending fashion in 0.5-degree steps until the temperatures required to evoke low, medium and high pain sensations could be reliably established. Three subjective pain levels were selected for each participant, corresponded to ratings of 2, 5 and 8 respectively on a continuous VAS scale with 10 numbered anchor points, where 0 represented “no pain” and 10 represented “unbearable pain.” Participants were instructed to rate undetectable stimuli with “no sensation” with 0, “noticeable, nonpainful sensation” with 1, “nonpainful warmth” with 2, “barely painful” stimuli with 3, “moderately painful” stimuli with 5, and stimuli “near the maximum you could tolerate” with 8. Ratings of 10 resulted in termination of the stimulation by the experimenter. The temperatures required to achieve consistent VAS ratings of 2 were used then used as the “warm” stimuli, while those required to achieve consistent VAS ratings of 8 were used as “hot” stimuli for the remainder of the experiment. Thus, temperatures were selected so that the reported subjective pain was approximately equivalent for all participants.

Thermal stimuli and trials. Stimuli were applied using a Medoc TSA-II Thermosensory Analyzer with a 1.5 x 1.5 cm Peltier thermode device. Stimulation epochs included 3-s ramp-up and 4.5 s ramp-down periods, with 17 s at target temperature, for a total of 24.5 s per stimulus. Each stimulus was presented in the context of a trial involving anticipation, stimulation, rating, and an inter-trial interval. The purpose of the anticipation period was for consistency with previous studies and to stabilize attention to nociceptive processing across trials. An auditory warning cue (~650 Hz sinusoidal tone) of 1 s duration cued the start of the anticipation interval, which persisted 4 s after the offset of the tone. The thermal stimulus followed at the end of this interval. At the end of the 24.5 s stimulation, another auditory cue (1 s duration) signaled participants to rate painful stimuli on the 10-point VAS scale. A 30.5 s inter-trial interval followed, which prevented habituation or sensitization, which can occur with shorter rest intervals. Independent pilot testing was used to establish timing parameters that were safe and did not produce sensitization, and the stability of pain ratings across trials was verified during the experiment, as shown in Fig. 6C.

Expectancy manipulation phase. During this phase, which was conducted prior to scanning, subjects were instructed that the painful Level 8 stimuli would be applied 5 times to each region, but that the analgesic would block pain on the placebo-treated region. In fact, Level 8 stimuli were applied to the control-treated region, and Level 5 stimuli were applied to the placebo-treated region. Skin regions tested did not overlap between manipulation and testing.

MRI Acquisition

Axial spoiled-gradient recall IR-Prep structural MRI scans were acquired in all subjects on either a 1.5 or a 3 Tesla scanner (Signa LX, General Electric, Milwaukee, WI) for anatomical localization of PET data (TE = 5.5, TR = 14, TI = 300, flip angle = 20^{o}, NEX = 1, 102 contiguous images, 256 matrix, 0.86 x 0.86 x 1.4 mm).

PET Acquisition and Reconstruction

PET scans were acquired with a Siemens HR^{+} scanner in 3-D mode (reconstructed FWHM resolution ~5.5 mm in-plane and 5.0 mm axially), with septa retracted and scatter correction. Participants were positioned in the PET scanner gantry, and two intravenous (antecubital) lines were placed in the right volar forearm. A light forehead restraint was used to eliminate intrascan head movement. [^{11}C]carfentanil was synthesized at high specific activity (> 2000 Ci/mmol) by the reaction of [^{11}C]methyliodide and a nonmethyl precursor as previously described (2), with minor modifications to improve its synthetic yield (3); 10-15 mCi (370-555 MBq) were administered during the scan. Radiotracer administrations for the two scans were separated by at least 2 hours to allow for radiotracer decay, and to eliminate any possible residual effects from the previous challenge.

The total mass of carfentanil injected was 0.026 ± 0.01 µg/kg per scan, ensuring that the compound was administered in tracer quantities, i.e., subpharmacological doses.Fifty percent of the [^{11}C]carfentanil dose was administered as a bolus, and the remainder as a continuous infusion using a computer-controlled pump to achieve steady-state tracer levels at ~35-40 minutes after tracer administration. Receptor occupancy by carfentanil is estimated to be between 0.2% and 0.6% across the brain (i.e., in regions with low, intermediate and high µ-opioid receptor concentrations), based on the mass of carfentanil administered and the known concentration of opioid receptors in the postmortem human brain (4, 5). For each scan, 28 frames of images were acquired over 90 min with an increasing duration (30 sec up to 10 min). Frames of

interest in analysis were collected during from 10-40 min for warm and from 50 to 80 min for hot thermal stimulation.

Reconstruction. Images were reconstructed using an iterative algorithm (brain mode; FORE/OSEM 4 iterations, 16 subsets; no smoothing) into a 128 x 128 pixel matrix in a 28.8 cm diameter field of view. Attenuation correction was performed through a 6-min transmission scan (^{68}Ge source) obtained prior to the PET study, also with iterative reconstruction of the blank/transmission data followed by segmentation of the attenuation image. Small head motions during emission scans were corrected by an automated computer algorithm for each subject before analysis, and the images co-registered to each other with the same software (6, 7). Time points were then decay-corrected during reconstruction of the PET data.

Quantification of binding. Reconstructed image data were transformed on a voxel-by-voxel basis into two sets of parametric maps: (a) a tracer transport measure (K_1 ratio), and (b) a receptor-related measure, distribution volume ratio (DVR, equal to binding potential (BP) + 1 or Bmax / Kd) + 1). To avoid the need for arterial blood sampling, the tracer transport and binding measures were calculated with a modified Logan graphical analysis (8), using the occipital cortex (an area devoid of μ-opioid receptors) as the reference region. The slope of the Logan plot was used for the estimation of the distribution volume ratio (DVR), a measure equal to the (Bmax/Kd) + 1 for this receptor site and radiotracer. Bmax/Kd (or DVR-1) is the receptor related measure (μ-opioid receptor availability, or binding potential). With the bolus-continuous infusion tracer administration method employed, the Logan plots become linear 5-7 min after the initiation of radiotracer administration, allowing for the calculation of DVR values early in the scanning periods (warm stimulation). K_1 and DVR images for each experimental period and MR images were co-registered to each other prior to further processing.]

PET-MRI image coregistration and spatial normalization

Because some brain structures of interest were of small size, particularly in the midbrain, and artifacts in group analysis are likely to result if images from different subjects are not closely aligned, we adopted the enhanced procedures described below for registering and standardizing brains across participants. Robust regression procedures (described below) further minimize the impact of local mis-alignment of one or a few participants' brains.

Coregistration. Coregistration procedures developed in our laboratory were employed to maximize the quality of the registration of MRI and PET images. MRI images were coregistered to the mean PET DVR image for each subject using an iterative procedure. Origins of MRI and the approximate origin PET were set at the anterior commissure manually, and gross rotation and translation of the MRI image was performed manually as needed prior to beginning the procedure. SPM2 software was then used for automated mutual information coregistration. As this process can result in a suboptimal, local-maximum solution, registrations were checked manually for alignment at 14 recognizable landmarks. The starting position of the MRI image was adjusted manually as needed, and the automated mutual information algorithm re-run until satisfactory alignment was achieved. Adequate alignment was achieved in most image sets after 1-2 iterations. In addition, registration quality was quantitatively compared across participants to identify those with sub-standard quality, as described below.

Spatial normalization. Nonlinear image-intensity based normalization of MRI images to the standard Montreal Neurologic Institute template (avg152T1.img) in SPM2 (discrete cosine transform basis set with 35 mm cutoff, low regularization, b-spline image

interpolation). We compared normalization to this template vs. the smoother avg305T1.img and T1.img templates and found that it produced superior results for this dataset (quantitative results described below were slightly better than using the same procedures in SPM5, so we elected to use SPM2). Deformations estimated from high-resolution MRI images were applied to PET images and resliced to 2 x 2 x 2 mm voxels.

To quantitatively evaluate normalization, we compared the similarity of normalized MRI images with the template image and with the group average according to several metrics: 1) Correlation coefficient in intensity values between the two images (mean = .60, standard error of the mean (SEM) = .03); 2) mutual information (64 bins; using the method used in SPM2), mean = .52, SEM = .04; 3) mean absolute intensity difference of Z-transformed (mean-centered and unit variance) images (mean = .45, STE = .01). Outlying subjects were identified (and re-normalized if necessary) by examining the distribution of these scores in the group and by visually comparing brains with the highest and lowest quality normalizations. These procedures resulted in acceptable normalization results for all participants.

Region of interest (ROI) definition and multiple comparison correction strategy

Because tests in broadly defined regions (such as "cingulate gyrus") involve many voxels and make it difficult to quantitatively control false positive rates, we electronically defined focused ROIs based on previous work. ROIs were sets of contiguous voxels that met all of the following criteria: 1) a priori interest based on previous placebo effects in fMRI or μ-opioid receptor BP; 2) high levels of μ-opioid receptor BP in the current sample (> 1.1, reflecting specific binding (see Fig. 6A); 3) activation within 10 mm in at least two previous studies of placebo, opioid administration, or emotion regulation, from the study set used in Fig 8 of Benedetti et al. (9). In some cases, ROI volumes were extended somewhat beyond the boundaries defined in criterion 3 above to cover the highest opioid-binding regions in the current sample. ROIs are overlaid on average opioid BP in Fig. 6B; the transparent brown-white color scale shows the same categories of BP as in Fig. 6A.

Eight primary ROIs included periaqueductal gray (PAG), rostral anterior cingulate (rACC), pregenual anterior cingulate (pgACC), medial orbital sulcus (MOS) in the mid-lateral orbitofrontal cortex, lateral orbitofrontal gyrus/inferior frontal gyrus (LOFC), amygdala, nucleus accumbens (NAC), and anterior insula (aINS). Each of these regions has been shown to be important for placebo effects in fMRI and/or opioid binding studies. These ROIs were grouped into 13 regions of contiguous voxels (some regions involved homologous structures in left and right hemispheres). Additional ROIs in the thalamus, dorsolateral prefrontal cortex (DLPFC), medial OFC, and dorsal caudate were also defined as above because of their role in opioids and pain regulation in previous studies, making 27 contiguous regions in all. Results are reported correcting for multiple comparisons within regions using small volume correction (SVC), and correction for multiple comparisons across ROIs was done using a set-level permutation test, as described below.

Correction for multiple comparisons. We corrected for multiple comparisons both within and across ROIs. We used nonparametric permutation tests to perform small-volume correction (SVC) for search over voxels within each a priori ROI. Correction across regions for the number of ROIs tested (at the 'set level') was also performed using the permutation test by comparing the number of significant SVC ROIs with the number expected under the null hypothesis. For ROI sets that show a significant number of activated regions, we report SVC-corrected regions (yellow in all figures). Details of the permutation test are described below.

An advantage of the set-level correction approach is that each ROI is tested individually, which is appropriate if each ROI is hypothesized to be active a priori. If more ROIs reach SVC significance than would be expected by chance in the omnibus set-level test, then the individual regions that show significant results may be meaningfully interpreted. Correction across all voxels, ignoring grouping into regions, will yield a less sensitive analysis because the more regions (and voxels) are reliably activated across studies (and thus the more ROIs), the less sensitive the analysis will be in all regions. Thus, widespread placebo effects will require many more participants to establish corrected significance than placebo effects in only limited regions. This undesirable feature is avoided in the set-level correction.

Statistical analysis

Statistical analysis was carried out in the General Linear Model (GLM) framework using iteratively reweighted least squares (IRLS). The IRLS GLM is contained in the Matlab R2006a function robustfit.m, and was tested and validated using simulations and on neuroimaging data (10). The Robust Regression Toolbox (written by TDW), freely available at http://www.columbia.edu/cu/psychology/tor/, contains the code used to run the analyses described in this paper.

GLM design matrix. The GLM design matrix consisted of an intercept, reported placebo analgesia for each subject (centered to be mean zero), and administration order (placebo first vs. control first, contrast coded). The estimate of the intercept is an estimate of group activation, controlling for reported analgesia and order. Fitting this model allowed us to test whether contrast estimates (e.g., PH – CH) were statistically different from zero for the group while simultaneously controlling for additional known sources of variance (reported placebo analgesia and order). Effects of reported placebo analgesia on contrasts in opioid binding are of theoretical interest and are presented in the results as brain-reported analgesia correlations. While order was controlled in all analyses, it was not a covariate of interest and is not reported.

Permutation and correction procedures for individual ROIs. IRLS multiple regressions were run on each voxel in each ROI. We used permutation tests to estimate the distribution of the maximum t-value in each ROI—and summary statistics including the number of individually significant ROIs in the whole set—under the null hypothesis of no placebo effects on opioid activity. This allowed us to correct for multiple spatially dependent tests without making any assumptions about the spatial covariance structure. IRLS regressions were repeated 10,000 times on permuted data. For each contrast of interest, the rows of the s x v data matrix (s = number of subjects x v = number of voxels in the analysis) were subjected to the same permutation on each iteration, preserving the spatial dependence in the data. The maximum t-statistic in each contiguous, a priori region of interest (ROI) was saved for each iteration, and the distribution of maximum t-values under the null hypothesis was thus estimated. Observed (correct permutation) t-statistic values within each ROI were thresholded at the 95^{th} percentile of the null hypothesis distribution, corresponding to $p < .05$ corrected for multiple comparisons (one tailed) or $p < .10$ corrected (two-tailed) within each ROI.

A different permutation method was appropriate for the intercept and for the covariates. What was permuted in each case is described in the next section.

Null hypotheses and permutation methods for group effect and covariates. For the intercept, the null hypothesis was a group average contrast value of 0, and the signs of the rows of the s x v data matrix were permuted (as in (11)) and t-values for the intercept estimated with IRLS, controlling for the covariates. For the covariates, a separate procedure was employed. The null hypothesis was a zero correlation between the data and each covariate of interest. The order of the rows of the s x v data matrix

were permuted, breaking the association between data and regressors while conditioning on the marginal distributions of both data and covariates. These two tests are discussed in (11), though we apply them here in the context of IRLS regression.

Pooling of weights in ROIs. Subject weights were pooled (averaged) within contiguous a priori regions of interest. Pooling of weights allows subject weights for a voxel to be influenced by those for neighboring voxels regularizing the distribution of weight values across space and making each subject's contribution to the group result the same throughout the ROI. This procedure is analogous to variance smoothing in the methods of Nichols (12) and Worsley (13), as it decreases the variability in the t-distribution estimates and increases sensitivity in the presence of multiple comparisons.

Set-level correction across ROIs. Even if ROIs are specified a priori and each is expected to show an effect of interest, there is potential for false positives if many ROIs are tested. Rather than raise the primary threshold and decrease the likelihood of finding true effects in ROIs, we report significance at small-volume corrected (SVC) levels within each ROI independently. To provide a test of whether there is significant activation in the set of ROIs as a whole, we used the permutation test to estimate the null hypothesis distribution of the number of significant ROIs that exceed the SVC threshold. Thus, if too few ROIs are significant, we cannot reject the null hypothesis that no effect is present in the set of ROIs as a whole. Alternatively, if more ROIs are significant than would be expected by chance, it can be concluded with confidence that one or more ROIs truly showed the effect of interest (e.g., placebo – control opioid activity).

Robust partial correlation coefficients. Though statistic values were obtained using IRLS GLM analysis, we report correlation effect sizes in terms of partial correlation coefficients between opioid activation and PRP. For this, subject weights from the IRLS procedure were used to determine a weighted correlation coefficient:

Weighted partial correlations do not reflect the inflation of error variance adopted in the IRLS procedure (according to (14)), and so are not always monotonically related to statistical significance. Thus, they are used descriptively in this report.

Multivariate analysis: nonmetric multidimensional scaling (NMDS) and clustering

Selection of regions. Thirty-two regions (groups of contiguous voxels) that showed placebo effects in one or more contrasts were selected for network analysis. These regions were all contained in the set of a priori ROIs and constitute contiguous groups of voxels within ROIs that were significant with either SVC correction or at $p < .005$ in one of the following contrasts: Placebo x Heat interaction, placebo effect during hot stimulation, placebo effect during warm stimulation. These regions constituted the set of voxels that showed placebo effects of some kind in our study. Sizes of regions ranged from 1 to 90 voxels, with a mean of 15.5 voxels.

Measure of connectivity. Connectivity between pairs of regions was estimated by correlating average opioid binding values (across warm/hot and placebo/control conditions) across participants. Spearman's ρ was chosen as a measure of connectivity because it is robust to violations of the normality assumption and less sensitive to outliers. It is slightly less sensitive than the standard Pearson's correlation, but due to the potential in neuroimaging studies of outliers arising from local variations in the quality of inter-subject registration and other sources, we feel that ρ provides estimates that more accurately represent the true underlying connectivity. Analyses with Pearson's r showed similar patterns of results and, as expected, were slightly more sensitive in many cases.

NMDS analysis. NMDS and its predecessor metric MDS are dimension reduction procedures, similar to principal components analysis (PCA), independent components analysis (ICA), and related techniques. The purpose of applying it to this dataset is twofold: first, to capture as much of the structure of connectivity as possible in a low-dimensional space that can be visualized in two or three dimensions; and second, to reduce the number of variables relative to the number of observations (participants, in this case) in subsequent cluster analyses. This dimension reduction reduces the sparseness in the clustering space and stabilizes the solution. The visualizations and clustering of regions based on patterns of opioid responses (Figs. 4, 11, and 12) uses the NMDS and clustering procedures described below. The tests of placebo effects on connectivity between pairs of regions do not depend on these procedures.

MDS analysis, like PCA, uses eigenvalue decomposition to find a set of canonical components that capture a maximal amount of variance in the data. Components are sorted by variance explained, so that the first components capture the most prominent patterns of systematic covariation among variables (in this case, variables are brain regions). Both MDS and PCA provide eigenvectors, or weights on brain regions for each component, and component scores. In this case, component scores are the canonical patterns of individual differences across participants. However, because MDS was developed for characterizing and visualizing patterns of similarity (or in this case, connectivity) in a set of variables (15), the method has a different theoretical emphasis, and additional measures beyond those used in classic MDS can increase the validity of inferences made on connectivity patterns.

The theoretical emphasis in MDS is on the similarity of brain regions, rather than on underlying distributed components that may contribute to multiple regions. The input data is a matrix of generalized dissimilarities (or distances) among a set of objects. The concept is that the pattern of dissimilarities can be thought of in terms of distances between objects (brain regions) in an underlying, unobserved multidimensional space. As a general example, consider that a 50 x 50 matrix of measured distances between cities is the product of a much simpler arrangement of cities in three-dimensional geographic space. The goal of MDS is to recover the locations in space from the pattern of similarities. If the measured distances are accurate enough, the 1225 observed inter-city distances in the 50 x 50 matrix can be perfectly represented as Euclidean distances among the 50 cities in three-dimensional space (only 150 parameters). An advantage is that all 1225 estimates contribute to the reconstruction of the objects in space so that error due to any single pairwise estimate is minimized.

In this case, the data is a matrix of correlations among brain regions transformed to represent dissimilarities using the formula: $d = (1 - \rho)/2$, where d is dissimilarity and ρ is Spearman's rho. Thus, ρ=1, a perfect correlation, yields = 0 (no dissimilarity), and ρ=-1, a perfect negative correlation, yields d = 1 (maximum dissimilarity). NMDS decomposes the matrix of dissimilarities D into a regions x dimensions set of stimulus coordinates for the regions in k-dimensional space. Brain regions closer together in this space are more similar in their profile of activation across subjects, and are more highly correlated. The dimensionality of the space is chosen to capture as much of the original similarity matrix with as few dimensions as possible. An error metric developed by Kruskal (16) and widely used is called stress, and is defined as follows:

$$stress = \sqrt{\frac{\sum_{i,j}(d_{ij} - \hat{d}_{ij})^2}{\sum_{i,j}\hat{d}_{ij}^{\,2}}}$$

where $\hat{d}_{ij}$ is defined as the Euclidean distance (the most commonly used distance metric) between regions i and j implied by the model, and d_{ij} are the observed dissimilarities between regions[1]. MDS algorithms minimize stress or a similar metric, and with one additional adjustment described below it is what is minimized by the NMDS algorithm we applied here. In the current analysis, an 8-dimensional space captured most of the variance in the data and so was deemed sufficient. With this dimensionality, the error is on average only about 3% as large as the average model-implied distance. Figure 11A shows the plot of stress (y-axis) against the number of dimensions included in the model (x-axis). The 'elbow' in the plot is the point at which including additional dimensions returns a lower payoff in variance explained, and the solution was chosen to be at the elbow in this function. Changing the dimensionality of the space has little effect on the solution, as the extra dimensions included or excluded have very little variance. However, including extra dimensions will have a destabilizing effect on the clustering solution described below, so is undesirable. Whereas the clustering solution and the visualization of regions in Figs. 4 and 12 depend on the dimensionality and choice of NMDS as a decomposition method, the tests of placebo effects on pairwise connectivity among regions does not depend on the NMDS or clustering.

Importantly, the classic MDS solution assumes that the model-implied dissimilarities ($\hat{D}$) in the reduced-dimensional model are linearly related to the measured dissimilarities. This is the case with cities, which are embedded in a truly Euclidean space, but it is not necessarily true for brain regions. If the observed dissimilarities are not linearly related to the distances in the assumed underlying similarity space, minimizing stress on Euclidean distances is not appropriate. The insight of Shepard (17, 18), however, was that it is not necessary to assume a Euclidean space—only to assume that observed dissimilarities are some monotonic function of the distances among regions in the underlying similarity space.
He found that from ordinal information about the relative dissimilarities alone it is very often possible to recover an underlying metric space.

Shepard proposed a check on whether the metric model is adequate: the observed dissimilarities are plotted against the model implied dissimilarities, and if the relationship is nonlinear, minimizing stress in a metric space is inadequate. The Shepard plot from our analysis is shown in Fig. 11B. The nonlinear relationship indicates that nonmetric MDS (NMDS) is more appropriate.

In NMDS, which was applied to this dataset, a nonmonotonic regression of the actual distances D against $\hat{D}$ is computed. Nonmonotonic (or isotonic) regression finds the best-fitting monotonically increasing function of D on $\hat{D}$. This is done by considering increasing values of D and aggregating (averaging) $\hat{D}$ values that are not monotonically increasing with previous values since the last monotonic increase. Successive values of $\hat{D}$ are averaged until the average is as high or higher than the previous values. If there is no relationship, the fit will approximate a flat line. The nonmonotonic regression plot is shown by the thin black line in Fig. 11B. The stress metric is adjusted to compare $\hat{D}$ with the best-fitting monotonic function of D, denoted $f(D)$:

[1] The subscript notation is used to indicate that measure sums squared deviations across all pairwise combinations of regions rather than across all elements of D.

$$stress = \sqrt{\frac{\sum_{i,j}(f(d_{ij}) - \hat{d}_{ij})^2}{\sum_{i,j}\hat{d}_{ij}^{\,2}}}$$

As there is no closed-form solution for minimizing nonmetric stress, the algorithm works iteratively by recomputing stimulus coordinates and assessing stress. The best fitting solution was found by gradient descent. To avoid convergence on a local minimum, we repeated the analysis 10 times with different random starting configurations chosen from a multivariate normal distribution and chose the best solution. This method produced highly replicable results across replications of the analysis (no visually observable differences were found in the NMDS plot in Fig. 4). The NMDS algorithm used was implemented in Matlab 7.2's (Mathworks, Natick, MA) mdscale.m.

Clustering of regions in component space. The goal of the analysis is to find sets of regions whose opioid binding levels are coherent across individuals (i.e., that show similar patterns of individual differences in binding levels) and distinguishable from that of other sets. Clustering algorithms are well-suited to finding such sets because they are designed to identify classes of nearby objects (here, brain regions).

By contrast, grouping regions that show similarly high loadings on individual PCA or ICA components is not an appropriate method for identifying sets of regions. This is because regions that load highly on one component are not necessarily very similar overall (they may differ on other components). In addition, critically, components in the PCA solution are determined up to rotation, which means that there is no inherent meaning to a region's loading highly on one component. The ICA solution is not computationally indeterminate, but the rotation of the component loadings is dominated by noise. Thus, as the entire solution may rotate in multidimensional space and produce identical (or near-identical in ICA) fits to the data, the set of regions loading on any one component may shift arbitrarily. Thus, the similarity of regions in multidimensional space rather than on any single component is an appropriate measure for grouping regions into coherent sets.

We used hierarchical clustering with average linkage (clusterdata.m in Matlab 7.2) to identify sets of regions. The NMDS component scores are estimated coordinates in the 8-dimensional solution space, and regions were grouped based on Euclidean distances in this space, (ignoring nonlinearities in the underlying space; this approximation may be improved on in future research). We used a permutation test to choose the number of clusters and to provide inferences on whether the distances between regions were truly distributed multimodally (as opposed to a single-mode, single-cluster distribution expected if there were no systematic sub-networks in the opioid system). For each possible solution between 2 and 19 clusters, we first computed a measure of clustering quality, as defined in (19):

$$q = \sum_k \sum_i \frac{d_{i_o} - d_{i_{nn}}}{\max(d_{i_o}, d_{i_{nn}})}$$

where d denotes Euclidean distance, and d_{i_o} is the distance from region i to the center of its own class, $d_{i_{nn}}$ is the distance to the nearest neighboring class, and k indexes over clusters. We then permuted the columns of the dimension scores, re-applied the clustering algorithm, and calculated *q* based on the permuted data. The permutation procedure disrupts clusters of nearby regions by exchanging their locations in each dimension with those of other regions, while conditionalizing on the marginal

distribution of regions in each dimension. This process was repeated 10,000 times to develop a null-hypothesis distribution of *q*. Estimating the distribution of q for each candidate number of clusters k allowed us to assess Z-scores observed-data clustering solution, defined as:

$$Z_k = \frac{q_{obs} - \overline{q}_{null}}{\sqrt{\frac{\sum_{1}^{I}(q_{null} - \overline{q}_{null})^2}{I}}}$$

q_{obs} is the quality for the observed solution, q_{null} is the quality for the permuted-data solution for one iteration, and I is the number of iterations (10,000). Fig. 11C shows Z_k on the y-axis plotted against candidate choices for k on the x-axis. The highest Z-values were found for 7 clusters, which we used as our estimate of k. The permuted-data distribution of q is shown for the 7-cluster solution in Fig. 11C, and q for the observed-data solution is shown by the vertical black line. The significance of the results ($p < .0001$, $Z = 4.0$) indicates that connectivity in opioid binding among regions was not unimodally distributed. Note that we did not statistically compare the 7-cluster solution against other candidate k's. Other choices of k may be reasonable candidates as well and statistically indistinguishable from the 7-cluster solution.

Comparing dependent correlations across conditions

Assessing whether the correlation between two regions is stronger under placebo than control conditions involves comparing dependent correlations. Because $\rho_{placebo}$ and $\rho_{control}$ are estimated on the same participants, the ρ values themselves are positively correlated. Steiger (20) provides several alternative statistics for estimating the covariance of dependent correlation estimates and constructing statistical tests. We adopt the recommended test statistic denoted $\overline{Z}_2^*$ (Eq. 14 of (20)).

To test whether placebo induced a shift towards positively intercorrelated regions among a set of regions (i.e., functional integration), we constructed a permutation test by permuting the rows of the participants x regions data matrices independently for each region. This procedure disrupts inter-region correlations present in the observed data while conditionalizing on the marginal distributions in all regions. The same permuted ordering was applied to placebo and control data matrices to ensure that a comparable manipulation was applied to each matrix. We counted the number of significant positive correlations minus significant negative pairwise correlations in the permuted dataset (both $p < .05$) to estimate the null hypothesis distribution of differences, and compared the observed-data differences in numbers of significant correlations against this null distribution to get p-values. Spatial dependence is not preserved in the permuted data, and this is a direction for future development.

Supplementary Results

Much of the supplementary results are contained in supplementary figures 6-12, below, and their respective legends. Additional notes are reported here.

Placebo effects in reported pain. Placebo treatment led to significant reported analgesia during noxious heat (PH – CH in reports, 5.58 for CH vs. 5.07 for PH, robust $t(13) = 1.87$, $p = .042$, one-tailed. The PW – CW contrast was not significant (1.57 vs. 1.42, $p > .10$), and the Temperature x Placebo interaction was marginally significant, $t(13) = 1.71$, $p = .055$, indicating greater placebo reduction during pain as expected.

Placebo administration order (placebo first vs. control first) did not interact with any of these effects. There was also no significant CH - PH effect of whether participants had participated in a previous behavioral session (0.81 for the returning group vs. .19 for the first-session group, t(9.6) = 1.04, p > .10). As planned prior to enrollment, opioid data were analyzed as a function of reported placebo analgesia, irrespective of returning-group status.

Functional integration with placebo

A remaining question is whether functional integration with placebo is specific to heat. Examination of inter-region correlations for placebo during warm stimulation revealed a significant integration PAG-rACC relationship in this contrast also, $\rho_{placebo}$ = 0.51 vs. $\rho_{control}$ = -.04, Z = 2.00, p = 0.04) suggesting that this functional integration may not depend on the level of noxious stimulation. Future studies with larger samples are needed address this issue more completely.

Causes of negative correlations

Negative correlations between reported placebo analgesia (C –P in pain report) and placebo increases in opioid activation (CH – PH and [CH – PH] – [CW – PW]) were found in many of the same areas that show group placebo-induced endogenous opioid activity increases in heat (CH – PH). Scatterplots are shown in Figs. 7 and 8; none seem to be driven by outliers, and robust regression techniques minimized the possibility of outlier-induced correlations. The finding of negative correlations in many regions is unexpected because they indicate that whereas opioid activation increases with placebo on average, high responders show less release. As mentioned in the main text, placebo responders also have lower BP across all conditions in many regions (see Discussion and Fig. 9). Supplementary voxel-wise analyses of correlations between opioid BP in control conditions (CW and CH) revealed widespread negative correlations with reported placebo in an around ROIs. Fig. 9D shows t-maps thresholded at p < .05 (two-tailed). Such correlations could occur for two reasons: Placebo responders may have higher tonic levels of endogenous opioids during testing, or placebo responders may have higher receptor binding affinity, which would lead to lower overall binding levels for a given endogenous opioid tone for these participants. Thus,

Responder vs. nonresponder differences in endogenous opioid tone could arise if responders release greater quantities of endogenous opioids in response to the experimental context or the manipulation phase preceding PET. If this is the case, the inclusion of the expectancy manipulation may be an important experimental difference from previous work (21). Alternatively, responders may have higher endogenous opioid tone for other reasons. Opioid tone has been assessed in rodents by measuring increases in glucose metabolism or c-fos expression after low-dose naloxone administration. Such studies suggest that there is tonic endogenous opioid activity in the brainstem and the amygdala of rodents that has a net inhibitory effect on metabolism (22, 23). In humans, naloxone administration has been shown to increase fMRI activity in a number of cortical (e.g., anterior cingulate, prefrontal and insular, entorhinal and parahippocampal cortices) and subcortical (e.g., basal ganglia, hippocampus) regions (24).

Whether placebo responder/nonresponder differences are related to endogenous opioid tone or individual differences in binding affinity, it remains to explain why high responders still reported less pain for [CH – PH] when opioid BP levels were less different across conditions. If low binding in placebo responders is caused by endogenous opioid release in CH, then a potential explanation could be that opioid

release has a larger analgesic effect when opioid levels are high than when they are low. If low binding is caused by individual differences in receptor affinity, then a unit increase in opioid release is expected to have a larger impact on pain in high-affinity participants (i.e., placebo responders). Both explanations rest on the idea that the relationship between opioid activity and pain is nonlinear—specifically, that, a unit increase in opioid release in subjects with low binding leads to a larger analgesic effect than a unit increase in subjects with high binding. This may be tested more directly in future studies with larger samples. The linearity of the relationship between opioid release and analgesia has not been extensively evaluated to our knowledge. However, if CH binding differences between high and low responders is related to differences in binding affinity, then a unit increase in release is expected to have a more potent effect in the high-affinity placebo responders; thus, a nonlinear relationship whose form is consistent with our data is expected.

A third class of explanations is that opioid release may be related to relief at the offset of stimulation. If so, low responders, who report more pain in the placebo condition, may experience more relief in that condition and thus show higher opioid levels. However, opioid release cannot be simply related to relief, because that hypothesis would predict that opioid activation should be higher in the control condition (more pain, more relief) than the placebo condition overall, which is the opposite of what we found. While a mixture of opioid release to pain and relief in the right proportions could conceivably explain both the main effects and correlations with reported placebo, this class of explanations does not fit the data as well. We tentatively favor the first explanation, of ceiling effects due to opioid release across all experimental conditions in high responders.

Supporting Discussion

A note on expectancy manipulation: Expectancy or conditioning?

While the expectancy manipulation we employ is similar to conditioning procedures used in some placebo studies (e.g., 25), it lacks a clearly defined unconditioned stimulus and response because no active treatment is being paired with sensory cues, and may operate via different mechanisms than classical conditioning. While the precise mechanisms remain ambiguous, our view is that the surreptitious reduction in temperature reinforces expectancies rather than creating specific conditioning in the classical sense.

A meaningful functional distinction between expectancy and conditioning might be made by defining conditioning in terms of specific, automatic learning of associations between a predictive sensory cue (a conditioned stimulus) and a target brain response (an unconditioned response) that occurs proximally in time (26), or a sensory event that elicits that response (an unconditioned stimulus). A delay of 500-1000 ms between conditioned and unconditioned stimuli may require different brain systems (i.e., hippocampus) and awareness (26), implying a different learning process. Expectancy, by contrast, involves deployment of conceptual knowledge to associate a context (which may have many features) with a pattern of responding (e.g., opioid release). The cue and response need not be associated proximally in time nor mediated by one specific neural pathway. A version of this distinction in the context of placebo effects is presented by Stewart-Williams and Podd (27). Conditioning may or may not elicit expectancies, but its defining feature is the specific nature of the associative learning and the temporal proximity required for this learning.

To consider our expectancy manipulation to be a conditioning procedure, at a minimum the conditioned stimulus and unconditioned response should be identified. The application of the placebo cream itself could act as a conditioned cue, but not one that would create a difference in pain reports: an identical cream is applied in the control condition, so the conditioned stimulus is the same for placebo and control conditions. What differentiates the conditions is the expectation, imparted verbally by the experimenter, that one cream has analgesic properties. A second variable that differentiates the placebo and control conditions is the history of sensory experience on placebo- and control-treated sites due to the expectancy manipulation. For this manipulation to induce conditioning, specific sensory cues unique to the placebo-treated site would have to be associated with an unconditioned response. Sensory information about the placement of the thermode on the placebo-treated skin could conceivably constitute such a cue. However, the response associated with those cues is one applied by the experimenter—the reduction of temperature—rather than any clearly identifiable response on the part of the participant. The unconditioned response could conceivably be endogenous opioid release related to the relief caused by the reduction of temperature, but this requires several speculative assumptions about the causes of opioid release with little data to support them.

Thus, because the expectation of analgesia is produced by changing the stimulus rather than by inducing an endogenous neurochemical response (as is likely in drug conditioning studies), the argument that our expectancy manipulation is a conditioning procedure is relatively weak. While we cannot rule out possible conditioning, our view is that the cognitive expectancy of reduced pain is primarily responsible for the placebo effects we observe, and that the expectancy manipulation reinforces these expectancies. Other research has shown that verbal suggestion alone produces opioid-dependent placebo effects (28, 29). In a telling study, verbal suggestions were pitted against presentation of stimuli designed to elicit conditioning (30). Expectancies elicited by verbal suggestions dominated potential conditioned effects, which were largely absent.

The following lists describe briefly each of the tables and figures in the Supporting Information. Table notes and figures legends are also provided with tables and figures.

List of Tables

List of Figures

References

1. Wager, T. D., Matre, D., & Casey, K. L. (2006) *Brain Behav Immun* **20,** 219-230.
2. Dannals, R. F., Ravert, H. T., Frost, J. J., Wilson, A. A., Burns, H. D., & Wagner, H. N., Jr. (1985) *Int J Appl Radiat Isot* **36,** 303-306.
3. Jewett, D. M. (2001) *Nucl Med Biol* **28,** 733-734.
4. Gabilondo, A. M., Meana, J. J., & Garcia-Sevilla, J. A. (1995) *Brain Res* **682,** 245-250.
5. Gross-Isseroff, R., Dillon, K. A., Israeli, M., & Biegon, A. (1990) *Brain Res* **530,** 312-316.
6. Minoshima, S., Koeppe, R. A., Mintun, M. A., Berger, K. L., Taylor, S. F., Frey, K. A., & Kuhl, D. E. (1993) *J Nucl Med* **34,** 322-329.
7. Minoshima, S., Maruno, H., Yui, N., Togawa, T., Kinoshita, F., Kubota, M., Berger, K. L., Uchida, Y., Uno, K., & Arimizu, N. (1993) *Ann Nucl Med* **7,** 71-77.
8. Logan, J., Fowler, J. S., Volkow, N. D., Wang, G. J., Ding, Y. S., & Alexoff, D. L. (1996) *J Cereb Blood Flow Metab* **16,** 834-840.
9. Benedetti, F., Mayberg, H. S., Wager, T. D., Stohler, C. S., & Zubieta, J. K. (2005) *J Neurosci* **25,** 10390-10402.
10. Wager, T. D., Keller, M. C., Lacey, S. C., & Jonides, J. (2005) *Neuroimage* **26,** 99-113.
11. Nichols, T. E. & Holmes, A. P. (2002) *Hum Brain Mapp* **15,** 1-25.
12. Nichols, T. & Hayasaka, S. (2003) *Stat Methods Med Res* **12,** 419-446.
13. Worsley, K. J. (2005) *Neuroimage* **26,** 635-641.
14. DuMouchel, W. H. & O'Brien, F. L. (1989) in *Computer Science and Statistics: Proceedings of the 21st Symposium on the Interface, American Statistical Association*.
15. Torgerson, W. S. (1952) *Psychometrika* **17,** 401-419.
16. Kruskal, J. B. (1964) *Psychometrika* **29,** 1-27.
17. Shepard, R. N. (1962) *Psychometrika* **27,** 125-140.
18. Shepard, R. N. (1980) *Science* **210,** 390-398.
19. Struyf, A., Hubert, M., & Rousseeuw, P. (1996) *Journal of Statistical Software* **1,** 1–30.
20. Steiger, J. H. (1980) *Psychological Bulletin* **87,** 245-251.
21. Zubieta, J. K., Bueller, J. A., Jackson, L. R., Scott, D. J., Xu, Y., Koeppe, R. A., Nichols, T. E., & Stohler, C. S. (2005) *J Neurosci* **25,** 7754-7762.
22. Gestreau, C., Le Guen, S., & Besson, J. M. (2000) *The Journal of comparative neurology* **427,** 285-301.
23. Kraus, M. A., Piper, J. M., & Kornetsky, C. (1996) *Brain Res* **724,** 33-40.
24. Borras, M. C., Becerra, L., Ploghaus, A., Gostic, J. M., DaSilva, A., Gonzalez, R. G., & Borsook, D. (2004) *J Neurophysiol* **91,** 2723-2733.
25. Amanzio, M. & Benedetti, F. (1999) *J Neurosci* **19,** 484-494.
26. Clark, R. & Squire, L. (1998) *Science* **280,** 77-81.
27. Stewart-Williams, S. & Podd, J. (2004).
28. Benedetti, F., Pollo, A., Lopiano, L., Lanotte, M., Vighetti, S., & Rainero, I. (2003) *J Neurosci* **23,** 4315-4323.
29. Benedetti, F., Arduino, C., & Amanzio, M. (1999) *J Neurosci* **19,** 3639-3648.
30. Montgomery, G. H. & Kirsch, I. (1997) *Pain* **72,** 107-113.

31. Duvernoy, H. M. (1995) *The human brain stem and cerebellum* (Springer-Verlag, New York).
32. Zambreanu, L., Wise, R. G., Brooks, J. C. W., Iannetti, G. D., & Tracey, I. (2005) *Pain* **114,** 397-407.
33. Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., Friston, K. J., & Frackowiak, R. S. (2004) *Nature* **429,** 664-667.
34. Keltner, J. R., Furst, A., Fan, C., Redfern, R., Inglis, B., & Fields, H. L. (2006) *J Neurosci* **26,** 4437-4443.