



# Sex-Specific Pharmacotherapy for Back Pain: A Proof-of-Concept Randomized Trial

Diane Reckziegel · Pascal Tétreault · Mariam Ghantous · Kenta Wakaizumi · Bogdan Petre · Lejian Huang · Rami Jabakhanji · Taha Abdullah · Etienne Vachon-Preseau · Sara Berger · Alexis Baria · James W. Griffith · Marwan N. Baliki · Thomas J. Schnitzer · A. Vania Apkarian

Received: May 14, 2021 / Accepted: July 15, 2021 / Published online: August 10, 2021  
© The Author(s) 2021

## ABSTRACT

**Background:** Preventing transition to chronic back pain (CBP) is a long-sought strategy that could rescue patients from prolonged suffering. Recent rodent and human brain imaging studies suggest involvement of sexually dimorphic, dopaminergic-motivational, mesolimbic circuits in the transition to chronic pain (tCBP), and hint that the combination of carbidopa/

levodopa and naproxen (LDP + NPX) may block tCBP. Here we evaluated, in people with recent-onset back pain, whether a 3-month treatment with LDP + NPX is safe, blocks tCBP, and whether its efficacy is sex-dependent.

**Methods:** A total of 72 participants were enrolled and stratified by risk for tCBP using brain-imaging biomarkers. Low-risk participants entered a no-treatment arm. Others were randomized to placebo + naproxen or LDP + NPX for 3 months.

**Results:** Both treatments resulted in more than 50% pain relief for approximately 75% of participants. A strong sex by treatment interaction

---

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40122-021-00297-2>.

---

D. Reckziegel · K. Wakaizumi · L. Huang · R. Jabakhanji · J. W. Griffith · M. N. Baliki · T. J. Schnitzer · A. V. Apkarian  
Center for Chronic Pain and Drug Abuse, Northwestern University Feinberg School of Medicine, Chicago, USA

D. Reckziegel · P. Tétreault · M. Ghantous · B. Petre · L. Huang · R. Jabakhanji · T. Abdullah · E. Vachon-Preseau · S. Berger · A. Baria · A. V. Apkarian  
Department of Physiology, Northwestern University Feinberg School of Medicine, Chicago, USA

K. Wakaizumi · R. Jabakhanji · M. N. Baliki  
Shirley Ryan AbilityLab, Chicago, USA

J. W. Griffith  
Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, USA

T. J. Schnitzer  
Department of Rheumatology, Northwestern University, Feinberg School of Medicine, Chicago, USA

T. J. Schnitzer · A. V. Apkarian  
Department of Physical Medicine and Rehabilitation, Northwestern University Feinberg School of Medicine, Chicago, USA

A. V. Apkarian  
Department of Anesthesia, Northwestern University Feinberg School of Medicine, Chicago, USA

A. V. Apkarian (✉)  
Physiology, Anesthesia, PM&R, Center for Translational Pain Research, Center of Excellence for Chronic Pain and Drug Abuse Research, Northwestern University, Feinberg School of Medicine, Tarry Bldg. 7-705, Chicago, IL 60611, USA  
e-mail: a-apkarian@northwestern.edu

was observed for daily pain intensity (phone NRS,  $P = 0.007$ ), replicated on 4-week average pain (Pain/4w,  $P = 0.00001$ ), and in intent-to-treat analysis (Pain/4w,  $P = 0.000004$ ). Nucleus accumbens functional connectivity with medial prefrontal cortex, a predefined objective biomarker, showed sex dependence at baseline ( $P = 0.03$ ) and sex-by-treatment interaction effect 3 months after treatment cessation ( $P = 0.031$ ). Treatment modified the psychological profile of participants, and disrupted brain modeling-based predicted back pain intensity trajectories. Forty participants were queried 3.3 years from trial start; back pain ratings were similar between end of treatment and at 3.3 years ( $P = 0.62$ ), indicating persistence of relief for this duration.

**Conclusions:** These results provide the first evidence for preventing transition to chronic back pain using sex-specific pharmacotherapy. These provocative observations require confirmation in a larger study. ClinicalTrials.gov identifier: NCT01951105.

**Keywords:** Dopamine; Naproxen; Prevention; Brain imaging; Chronic pain

### Key Summary Points

Brain dopaminergic circuits are implicated in the transition to chronic pain both in human brain imaging and rodent model studies

Preventing transition to chronic pain would be an ideal strategy to reduce the burden of chronic pain

The study tested whether the transition to chronic back pain can be blocked using a novel combination therapy (LDP + NPX) as compared to control (placebo + NPX), whether the response is sex-dependent, and if brain imaging can provide confirmatory evidence

There was no difference in the prevention of transition to chronic pain in subjects treated with LDP + NPX, compared to placebo + NPX (primary endpoint)

In prespecified subgroup analysis, there was a marked and sustained pain reduction in female participants treated with LDP + NPX in comparison to placebo + NPX, and brain imaging results were confirmatory

These results suggest that chronic back pain may be prevented using sex-specific pharmacotherapy, and a larger and better controlled trial is warranted

## INTRODUCTION

Chronic pain dramatically diminishes the quality of everyday life in about 20% of the world population, and for 100 million American adults [1]. Since a large proportion of these patients are treated with opioids, chronic pain remains a primary contributor to the US opiate epidemic [2]. Once established, its reversal becomes very difficult. Available treatments for chronic pain do not cure the condition, and most patients remain dissatisfied with their treatments. Thus, there is universal consensus that new, non-opioid and efficacious treatments, especially for back pain, are urgently needed. Chronic low back pain (CBP) is a highly prevalent type of chronic pain, and its related neural mechanisms are now well characterized [3–7]. The current trial was limited to subjects with back pain persistent over multiple weeks, i.e., subacute back pain (SBP), as such patients have a high probability of developing CBP [5, 8, 9] and the duration of their pain is within a time window during which it may be more malleable and thus reversible with proper treatment. Treatments that could block the transition from subacute to chronic pain (tCBP) and maintain pain relief are an ideal preventative strategy to combat chronic pain, as they spare such patients from years of disability, diminish the probability of exposure to opiate treatments, and decrease the associated staggering healthcare cost.

Mesocorticolimbic (MCL) brain properties have been previously shown to predict tCBP

[5, 10, 7]. Complementarily, in rodent models of chronic pain, components of this circuit show a causal relationship with pain-like behaviors, specifically neuronal activity in the medial prefrontal cortex and nucleus accumbens [5, 11, 12]. Within nucleus accumbens, peripheral neuropathic injury is accompanied by a decreased concentration of dopamine, changes in the excitability of dopamine D2 receptor-expressing spiny neurons, shrinkage of dendrites and decreased synaptic inputs [12], as well as diminished dopamine D1 and D2 receptor expression [13]. When such animals are treated with a combination of levodopa/carbidopa and naproxen (LDP + NPX) for multiple days, pain-like behaviors are diminished and many of the abnormalities of nucleus accumbens D2 spiny neurons are reversed [12]. Importantly, treatment with either compound separately was not effective. Additionally, this combination treatment was highly effective when administered within days after the peripheral injury but diminished in efficacy when administered at later times [7]. Therefore, here we test the concept that early aggressive treatment of SBP can prevent tCBP and maintain pain relief. Furthermore, most patients with chronic pain are women [14, 15, 16], and MCL shows evidence for sexual dimorphism [17]. In particular, dopamine release in nucleus accumbens is lower in women [18]. Therefore, we hypothesized that the efficacy of the combination pharmacotherapy in SBP would be sex-dependent.

Thus, the primary aim of this study (NCT01951105) was to test whether tCBP can be blocked using a novel combination therapy (LDP + NPX). Our main secondary aim was to test the sex dependence of response to LDP + NPX treatment. A third, exploratory aim, was to assess the duration of pain relief following treatment cessation. We also sought to explore the influence of treatment on the pain-related psychological profile, and additionally tested whether neuroimaging-related biomarkers could provide objective evidence regarding treatment efficacy.

## MATERIALS AND METHODS

This 24-week double-blind parallel randomized controlled trial was conducted at Northwestern University (Chicago, USA). Protocol and informed consent forms were approved by Northwestern University IRB as well as National Institute of Dental and Craniofacial Research/National Institutes of Health (NIDCR/NIH). All enrolled participants provided written informed consent. Safety and trial oversight were monitored by an independent data safety monitoring board and a clinical research organization, with NIDCR/NIH oversight. The trial was registered on ClinicalTrials.gov, under registry NCT01951105.

### Participants

Individuals with a recent onset of low back pain (SBP) were recruited through online social media, local advertising, and via Northwestern Medicine Enterprise Data Warehouse. Criteria for enrollment included history of low back pain with time from onset between 4 and 20 weeks, with signs or symptoms of radiculopathy, average reported pain intensity greater than 4 (on a numerical rating scale (NRS) [19] from 0 to 10) on the week before baseline assessment and the week preceding treatment start. Subjects were excluded if they reported other chronic painful conditions, systemic disease, history of head injury, psychiatric diseases, or more than mild depression (score > 19, according to Beck's Depression Inventory [20]). See Supplementary Material Note S2 for a full description of the inclusion and exclusion criteria.

### Trial Design

Eligible subjects first underwent a brain imaging session and, using predefined brain-derived parameters, were stratified into high- and low-risk subgroups for transitioning to CBP (see below). Those in the low-risk category, who had 60% or higher probability of recovering, entered the NoTx arm; the rest, who had less than 60% probability of recovering naturally, were

randomized between a placebo control arm and a pharmacological treatment arm. A flexible dose escalation procedure was used and both researchers and participants were blinded to treatment type. Study medications were tapered off at week 12, and participants continued to be evaluated for another 12 weeks. At this final visit, participants underwent a second brain scan. Figure 1a shows the study timeline and the types of data collected.

The design was set up to evaluate safety, efficacy, and sex dependence of the combination of carbidopa/levodopa (12.5 mg/50 mg–25 mg/100 mg–50 mg/200 mg, dose escalation based on response) plus naproxen 250 mg (LDP + NPX) administered three times a day, compared to placebo plus naproxen 250 mg (PLC + NPX) administered three times a day.

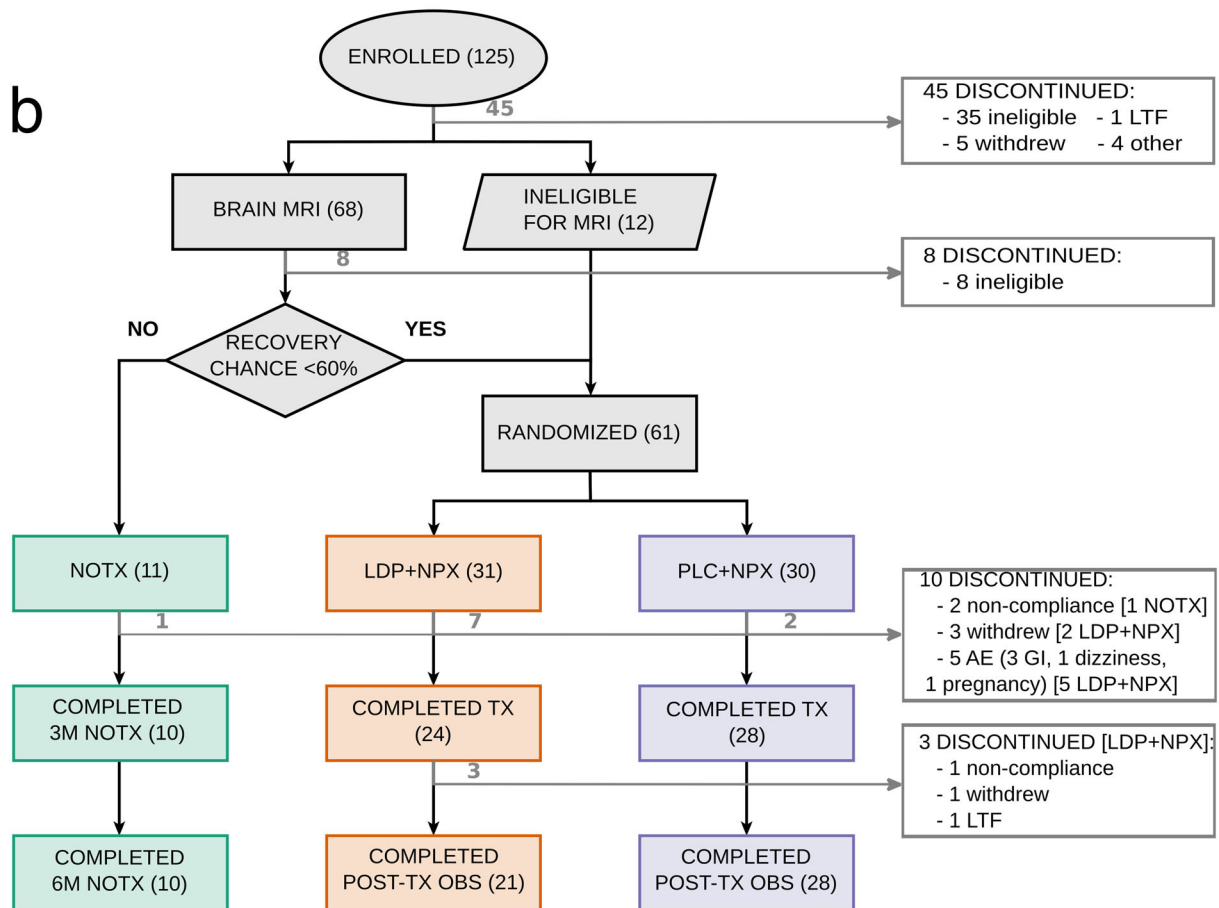
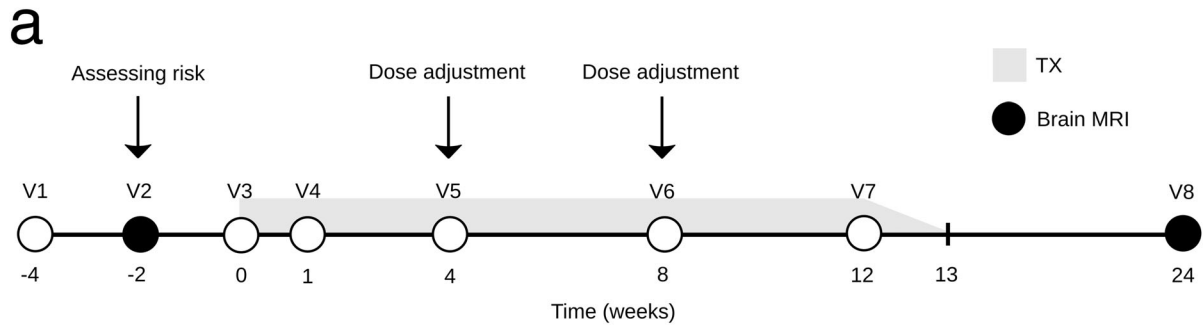
### Modeling Risk for Chronic Pain and Stratifying Participants

We used a naïve Bayes classifier to estimate probability of recovery from back pain for each participant before they entered the treatment phase. The classifier was trained using data from our previous longitudinal study in SBP and employed two brain markers to predict risk for CBP. These were the mean fractional anisotropy (FA) of white matter regions of interest that predicted persistence of low back pain at 1 year in previous work [10], and the number of functional connectivity links (degree count,  $D_e$ ) between the right anterior hippocampus, and limbic and pain-related areas (Supplementary Material Note S3). Every participant eligible for MRI underwent an initial brain scan, and brain-derived FA and  $D_e$  measures were used to identify probability for recovery. This measure was used to stratify participants into treatment and no-treatment arms. Participants who could not be scanned ( $n = 12$ ) were assumed to belong to the persisting category (as expected incidence rate for persisting category was approximately 80%) and were entered into the treatment arm. Model calculation details are described in Supplementary Material Note 4.

### Randomization and Masking

Researchers and participants were blinded to treatment allocation. Eligible individuals were assigned to treatment arms (LDP + NPX, PLC + NPX) on the basis of a computer-generated permuted block randomization scheme, with block size randomly varying and an allocation ratio of 1:1. Allocation concealment was ensured by utilization of sequentially numbered containers. An unblinded individual from Northwestern University Clinical and Translational Sciences (NUCATS), with no other role in the study, was responsible for ensuring proper medication assignment to each container. The randomization code was maintained by NUCATS and was available in cases of emergency or clinical situations in which knowing the treatment allocation would make a difference in the safety or management of a participant's health. In such a circumstance, the allocation assignment was made available after consultation with the site investigator and the principal investigator. At study conclusion, after database lock, the randomization code was made available to researchers analyzing the data, but only after all data preprocessing was completed and only when ready to perform statistical modeling.

Participants received two capsules on a three times a day (TID) schedule and one capsule once a day (QD). TID medications included one capsule of naproxen and one capsule of either placebo or some dose of carbidopa/levodopa. Omeprazole 40 g was taken QD in the morning, as a preventive measure against the adverse gastric effects of naproxen. In order to ensure that participants took their study medication as designed, the naproxen and omeprazole were placed in a separate-colored capsule from the carbidopa/levodopa. Each colored capsule was dispensed in separate containers and participants were asked to take one capsule from each container. Placebo capsules were identical to those containing LDP. Acetaminophen was available as a rescue medication and all participants were given equal amounts.



◀**Fig. 1** Trial profile. The study was a 28-week (8 visits, V1–8), randomized, 3-arm, flexible-dose, parallel-group trial of pharmacological treatment. **a** After a 2-week screening period (V1–2), brain MRI at V2 was collected and used to calculate the probability of recovery from back pain, in the absence of treatment. The V2 MRI result dictated whether a participant entered the randomization arm for treatment (receiving either LDP + NPX or PLC + NPX) or the no-treatment (NoTx) observation-only arm. Medications were administered for 12 weeks (V3–7), dose adjustments were made at V5 and V6, and tapered over weeks 12–13. All participants were followed for an additional 12 weeks (V7–8) to evaluate the persistence of benefit at the endpoint, 24 weeks after randomization and 12 weeks after treatment cessation. Questionnaires were administered at all visits. Adverse events assessments were performed at every visit. Blood screening was performed at baseline, while blood pressure, pulse, and temperature were assessed at all study visits. The V8 MRI was collected to assess treatment effects on predefined brain biomarkers. All patients reported daily back pain intensity, using a smartphone app. **b** Diagram shows numbers of subjects who met criteria for inclusion in, or exclusion from, the study, and their distribution among the three study arms. Entry in the treatment groups required a probability of recovery lower than 60% (determined between V2 and V3), according to a prediction model based on baseline brain scans. LDP + NPX levodopa/carbidopa + naproxen, PLC + NPX placebo + naproxen, TX treatment, NoTx no-treatment, OBS observation, AE adverse event, GI gastrointestinal distress, LTF lost to follow-up

## Procedures

The study consisted of eight visits spread over approximately 28 weeks, which included an initial screening/monitoring period (ca. 2 weeks), followed by a treatment period (12 weeks) and a post-treatment observational period (12 weeks) (Fig. 1a). Intermediate visits allowed for safety and clinical monitoring of participants, as well as dose adjustments. Safety and adverse events were assessed at each visit while participants were receiving medication (four visits), including measuring vital signs (Fig. 1a). After treatment cessation, adverse events were monitored via two phone calls. At each visit, researchers reminded participants to

take their medication according to instructions. Safety was also assessed through blood screenings at baseline, while blood pressure, pulse, and temperature were assessed at all study visits.

We considered both adverse events (AEs) that occurred during treatment and those that occurred after treatment. To account for potential late-onset AEs or those related to withdrawal of medication, we considered AEs occurring up to 28 days after the last dose of study drug, or end of study participation, whichever occurred first.

## Dose Escalation Protocol

Treatment with carbidopa/levodopa was titrated up to 12.5 mg/50 mg three times/day over 1 week and then continued at that level for 4 weeks. If by the end of the initial 4-week period the participant “responded” [had a greater than 20% decrease in pain intensity from the average of all phone NRS ratings collected at baseline (between visit 1 and visit 3) to the average of all phone NRS ratings during those 5 weeks], the participant was maintained on that dose for the duration of the treatment period (12 weeks total). If there was no response, the carbidopa/levodopa dose was increased to 25 mg/100 mg three times/day for the following 4 weeks, at which time the pain status was re-evaluated (based on the average of all phone NRS rating during those 4 weeks, against baseline average). Again, if a response occurred, that dose was maintained in a blinded manner for the following 4 weeks of treatment; if not, further dose titration of carbidopa/levodopa occurred to 50 mg/200 mg three times/day for the final 4 weeks. When a participant experienced an adverse event at higher doses, the participant was given the next lower dose that they were able to tolerate and then maintained on that dose for the remainder of the 12-week dosing period. Naproxen dose (250 mg three times/day) remained constant for all participants throughout the study, except it was not given during the tapering down at the end of the study. Both LDP + NPX and PLC + NPX arms were identically subjected to this dose escalation protocol. The number of capsules

taken per day was kept constant throughout the study, regardless of dose escalation.

### Monitoring Pain Intensity with Phone App

The main outcome measure was a numerical rating scale (NRS) [19], where “0” corresponds to no pain and “10” indicates worst possible pain. Participants were instructed to provide such ratings three times a day for at least 1 week prior to randomization, and throughout study participation, via a smartphone app (phone NRS), which contained the following instructions: “Please rate your current level of pain”.

### Phone NRS Preprocessing

Percentage residual pain values at 3 and 6 months were computed on the basis of the average phone NRS score from the week preceding treatment and the final week of treatment or post-treatment.

For each subject, percentage residual back pain (% Residual pain<sub>*i*</sub>) at a given day *i* was computed as:

$$\% \text{ Residual pain}_i = 100 - 100 \times [( \text{pain}_b - \text{pain}_i ) / \text{pain}_b],$$

where pain<sub>*b*</sub> is the baseline pain intensity, pre-defined as the average daily phone NRS during the week preceding the start of treatment, and pain<sub>*i*</sub> is the mean phone NRS at day *i*.

Thus, 100% residual pain reflected no change from baseline levels, while 0% residual pain reflected complete recovery. Participants were deemed responders if residual pain at 6 months was 80% or less (representing 20% pain relief), as defined a priori per protocol.

Specifically, when an individual had more than three ratings in a given day, we only kept three: the first, the second, and the last rating. Next, the average of these ratings was computed for each day. For missing daily ratings, the mean from nearest neighbors were used to replace the missing value. Two subjects who completed the study were not included in the

phone NRS analyses because of non-compliance in providing phone ratings.

We also examined outcomes at 3 months (i.e., at end of treatment), which was defined as the average phone NRS residual pain during the final week of treatment. For analyses of pain intensity trajectories, we included 7 days preceding treatment, 78 days during treatment, and 73 days post-treatment.

### Other Outcomes

In addition to phone ratings, participants also rated their current pain intensity (NRS) and were administered the following self-report questionnaires during each in-person visit:

1. Pain Sensitivity Questionnaire (PSQ) [21], a 17-item instrument used to assess individual pain sensitivity—it is based on pain intensity ratings of hypothetical situations, which includes various modalities (heat, cold, pressure, pinprick) and measures (pain threshold, intensity ratings). It can be split into two subscales: one consisting of items referring to mildly painful situations (minor, PSQ/min), and one consisting of the items referring to moderately painful situations (moderate, PSQ/mod).
2. Pain Disability Index (PDI) [22], an assessment of physical impairment in relation to pain.
3. PainDETECT [23], a 12-item assessment of neuropathic-like symptoms. PDI includes questions of current pain intensity (Pain/c) and a subjective report of average pain intensity over the past 4 weeks (Pain/4w).
4. McGill Pain Questionnaire—Short Form (sf-MPQ) [24], a well-validated measure assessing both sensory and affective components of pain (MPQ/s and MPQ/a). It also includes a visual analog scale (VAS) of pain.
5. Pain Catastrophizing Scale (PCS) [25], a 5-point instrument to assess 13 thoughts or feelings on past pain experience. PCS yields three subscale scores assessing rumination (PCS/r), magnification (PCS/m), and helplessness (PCS/h).
6. Pain Anxiety Symptoms Scale (PASS) [26], measures fear and anxiety responses specific

to pain. PASS consists of four aspects of pain-related anxiety: cognitive suffering (PASS/c), escape-avoidance behaviors (PASS/e), fear of pain (PASS/f), and physiological symptoms of anxiety (PASS/a).

7. Beck Depression Inventory (BDI) [20], a 21-item instrument for measuring the severity of depression.
8. Positive and Negative Affect Scale (PANAS) [27], has two mood scales, one measuring positive affect and the other measuring negative affect (PANAS/n). Each scale is rated on a 5-point, 10-item scale.

These measures were considered secondary and were used to construct back pain-related psychological profiles and examine treatment effects globally.

Information on sex, race, and ethnicity were obtained using a standard National Institutes of Health demographics form. Although this form did not explicitly differentiate between gender and biological sex, we believe it has captured information on biological sex.

Candidate brain markers of treatment efficacy and sex dimorphism were predefined as the extent of information sharing, at rest, between three key regions of interest: the right nucleus accumbens, medial prefrontal cortex, and insula. Specifically, insula connectivity was hypothesized to reflect pain intensity, as found in a previous study [4], while mesolimbic connectivity was hypothesized to reflect sex specificity [17, 18]. Thus, for each participant, we computed the connectivity strength between each pair of seeds, yielding three measures from fMRI scans collected before and after treatment. We related these to our primary outcomes. No additional brain analyses were performed.

### Persistence of Pain Relief

After the study closed, we contacted all the participants to ask about the status of their back pain. Our primary intent was to assess whether the pain relief observed would revert to original levels, or if the observed pain relief was persistent. The outcome measure was 4-week pain (PainDETECT). We also administered additional

questions mainly to ensure that participants remembered the study.

### Brain Imaging Acquisition

Data were acquired on a clinical 3-T Siemens Magnetom Prisma whole body scanner equipped with a receive-only 64-channel head/neck coil. At both baseline and 6 months after randomization, participants underwent a high-resolution anatomical scan (T1-weighted MRI), one resting and one spontaneous pain rating functional MRI (fMRI), and a white matter fractional anisotropy (FA) assessment scan (diffusion-weighted MRI). The entire MRI scan consisted of approximately 35 min of actual image acquisition, plus around 25 min for setting patients comfortably in the scanner, attempting to minimize their back pain, and for reacquiring images in case of excessive motion.

#### *T1-Weighted MRI Acquisition*

High-resolution T1-weighted brain images were collected using integrated parallel imaging techniques (PAT; GRAPPA) and the following acquisition parameters: voxel size  $1 \times 1 \times 1 \text{ mm}^3$ , TR = 2.3 s, TE = 2.40 ms, TI = 900 ms, flip angle =  $9^\circ$ , 176 sagittal slices, and acceleration factor = 2. Phase encoding direction was anterior to posterior, and the duration of acquisition was approximately 5 min.

#### *Resting-State fMRI Acquisition*

Blood oxygen level-dependent (BOLD) T2\*-weighted multiband accelerated echo-planar images were acquired at rest. Acquisition parameters were as follows: TR = 555 ms, TE = 22 ms, flip angle =  $47^\circ$ , 64 slices acquired with interleaved ordering, FOV = 208 mm, matrix size =  $96 \times 104$ , spatial resolution  $2 \times 2 \times 2 \text{ mm}^3$ , acceleration factor = 8. Phase encoding direction was posterior to anterior. Slices were acquired with ascending order to preserve the continuity of connections. The acquisition lasted approximately 10 min, during which 1110 volumes were collected. Participants were instructed to keep their eyes open



and to remain as still as possible during acquisition.

### ***Spontaneous Pain Rating fMRI Acquisition***

Identical acquisition parameters and duration to resting-state fMRI were used to obtain BOLD T2\*-weighted images while participants used a finger-spanning device to continuously rate and log the rate of their spontaneous back pain on a scale of 0–100, in the absence of external stimulation. Participants were instructed to keep their eyes open and to remain as still as possible during the scan.

### ***Diffusion-Weighted MRI Acquisition***

Multi-slice echo planar imaging with multiband excitation and multiple receivers was used to obtain diffusion-weighted images along 30 and 64 evenly spaced and non-collinear directions, with weighting factors of  $700 \text{ s/mm}^2$  and  $2000 \text{ s/mm}^2$ , respectively. Two non-weighted volumes were acquired, each at the beginning of each scan. Voxel size =  $2 \times 2 \times 2 \text{ mm}^3$ , TR = 3.5 s, TE = 92 ms, FOV = 230 mm, matrix size = 116, number of slices = 72, flip angle =  $90^\circ$ , multiband factor = 3, acquisition time approximately 7 min.

### **Brain Imaging Processing (for Baseline Model Parameters)**

For each subject, MRI data was processed within less than a week from the baseline acquisition and before the randomization visit, in order to extract brain parameters for patient stratification. The quality of each image modality was assessed for excessive motion and poor signal to noise ratio before preprocessing.

### ***Spontaneous Pain Rating fMRI: Right Hippocampus Connections with Limbic and “Pain” Regions***

fMRI volumes were preprocessed using tools within the FMRIB Software Library 5.0.9 (FSL), and MATLAB R2016a. After removing the first 120 volumes of each spontaneous pain rating functional dataset for magnetic field

stabilization, skull extraction using BET and slice-time correction were performed. `fsl_motion_outliers` was used to remove the effect of intermediate to large motion. Next, the remaining 990 volumes were filtered with a band-pass temporal filter (using Butterworth;  $0.008 \text{ Hz} < f < 0.1 \text{ Hz}$ ) and a non-linear spatial filter (using SUSAN; FWHM = 5 mm). Finally, nine vectors were regressed out, including the six parameters obtained from intra-modal motion correction using MCFLIRT, global signal (averaged over all voxels of the brain, over the 990 volumes), white matter signal (eroded white matter mask), and cerebrospinal fluid signal (eroded ventricular mask). These nine vectors were filtered with the Butterworth band-pass filter before being regressed from the time series to avoid recontamination.

The cleaned images were then linearly registered to the MNI template using FLIRT and down-sampled to  $6 \times 6 \times 6 \text{ mm}^3$  voxel size. A C-based, in-house program called “ABLM” (Apkarian Brain Linkage Map), previously described by Baria et al. [28], was used to compute the mean count of functional connectivity links (degree) between voxels within a predefined mask within the right hippocampus and limbic and “pain”-related regions. Link density threshold for calculation of degree was set to the top 10% of connections.

### ***Diffusion MRI: Mean FA***

Preprocessing of diffusion-weighted images was performed using eddy current to correct for eddy current-induced distortions and subject movement. DTIFIT was used to estimate the diffusion tensor in each voxel by linear regression and FA maps were derived. Following this, each FA map was non-linearly registered to the FMRIB58\_FA template. Next, we extracted the mean FA value of a group of voxels that was previously identified as having predictive value for transition to chronic pain [10] and used this value in our naïve Bayes model for stratifying participants between high- and low-risk SBP (further details about the model are given below).

## Brain Imaging Processing (for Post Hoc Analyses)

For post hoc analyses of brain function, we used our most recent preprocessing pipeline, which was optimized for longitudinal investigation. After removal of the first 20 volumes (corresponding to 11 s) of each functional dataset for magnetic field stabilization, data were subjected to skull extraction using BET, slice-time correction, motion correction with MCFLIRT, intensity normalization, and high-pass temporal filtering (0.008 Hz). Motion censoring was next performed by detecting volumes with framewise displacement (a measure of how much the head changed position from one frame to the next) larger than 1 mm, DVARS (indexes the rate of change of BOLD signal across the entire brain at each frame of data) with z score larger than 2.3, or BOLD signal z score  $> 2.3$ , and removing their adjacent volumes (-5 -4 -3 -2 -1 0 1 2 3 4 5). Signals from three vectors were regressed out, including global signal (averaged signal over all voxels of the brain, over the 1090 volumes), white matter signal (extracted from an eroded white matter mask), and cerebrospinal fluid signal (extracted from an eroded ventricular mask). Finally, the cleaned time series were band-pass filtered (0.008–0.1 Hz) to keep the low-frequency fluctuations of interest.

Functional image registration was optimized for longitudinal analysis by utilizing a two-step approach that minimizes within-subject variability. First, for each subject, functional images from scan 1 and scan 2 were registered to each other using forward and backward halfway linear transformations (FLIRT, six degrees of freedom). This process allows optimal within-subject alignment, with minimal displacement artifacts for both scans [29]. Second, the midway template was registered to MNI152 space using non-linear registration (FNIRT). In order to minimize warping effects, all non-linear registrations were constrained by affine registration. Final images were spatially smoothed with a Gaussian kernel (2 mm sigma). Registered data were visually inspected to ensure optimal alignment.

## Statistical Considerations and Analyses

All subjects who completed the study ( $n = 59$ ) were included in primary analyses (PLC + NPX,  $n = 28$ ; LDP + NPX,  $n = 21$ ; NoTx,  $n = 10$ ); all enrolled subjects with valid data ( $n = 71$ ) were included in intent-to-treat analyses (PLC + NPX,  $n = 30$ ; LDP + NPX,  $n = 30$ ; NoTx,  $n = 11$ ).

The predefined primary outcome measure for efficacy, as a function of treatment type and relative to sex, was the percentage of participants recovering from back pain. We used Fisher's exact test to test response efficacy as a function of type of treatment. Additionally, we employed ANOVA, as predefined in the protocol, for analysis of repeated measures of pain intensity trajectories. Given that this approach may lead to inflated degrees of freedom, we reanalyzed these data using two more conservative approaches (described below), both of which confirmed our original observation. The main secondary outcome was the sex dependency of treatment effects.

Other secondary endpoints were to test the validity of the model used for stratifying SBP and examine changes in brain connectivity with treatment. Exploratory analyses were conducted to examine treatment effects on pain-related questionnaire subscales, using dimensionality reduction methods and network analyses [30]. Reported binary outcomes are based on absolute and relative descriptive statistics, consistent with CONSORT guidelines. All statistical tests were two sided.

### *Confirmatory Time-Series Model 1*

To assess the presence of a treatment by sex and time interaction, while taking into account that points are nested within participants and are correlated in time, we constructed a hierarchical linear model with treatment, sex, and time (and their interactions) as predictors, and random intercepts and random slopes for time were permitted. Before fitting the model, we added 1 to the residual pain score (to ensure finite values) and log-transformed the resulting values; this served as the response variable. The resulting model is equivalent to an exponential decay

but is linearized for numerical stability. We fit the model using the lme4 package in R, and *P* values were calculated via Satterthwaite degrees of freedom, estimated using the lmerTest package in R.

### ***Confirmatory Non-parametric Time-Series Model 2***

We performed independent ANOVAs to investigate the treatment by sex interaction over time, comparing each time point time to the baseline *t*<sub>0</sub>. This approach makes no assumptions about the temporal correlation.

### ***Confirmatory Modified Intent-to-Treat Analysis***

To account for all individuals who received at least one dose of study medications, we carried out an intent-to-treat analysis, using imputation of outcome measures to 6 months from start of trial. Specifically, scores from the last time-point for which data was collected were imputed to 6 months. One person admitted answering questionnaires and pain ratings in a random manner, and therefore was not included in any of the analyses.

### ***Power Analysis for Primary Outcomes***

Because our primary dependent variable was binary (recovery from back pain or not), and to have sufficient power to detect treatment effects, we conducted power analysis based on comparison of independent proportions. Our previous study in SBP shows that in high-risk SBP approximately 90% persisted with back pain if treated with standard of care (mostly occasional use of anti-inflammatories) [5]. Thus, for power calculations, we assumed that 90% of participants that receive placebo and naproxen would have persisting pain at study end. Calculations in G\*Power 3.1.9.2 indicated that we would detect significant differences with 80% power if the probability of pain persistence is reduced to 67% or smaller (type I error rate assumed at 0.05, two-tailed, *n* = 50 per group). We assumed a two-tailed comparison, even though there was no a priori reason for the placebo plus naproxen to outperform the test drug. We therefore planned to enter 126

participants, accounting for attrition and for the no-treatment group.

Our second aim was to determine if there is a sex by treatment interaction effect. Specifically, we wanted to test the influence of sex on response to active treatment with carbidopa/levodopa and naproxen. Assuming no sex effect on the placebo and naproxen treatment response rate, we are left with the active treatment arm. Calculations in G\*Power 3.1.9.2 indicated that we would detect an odds ratio greater than 2.7 (2.5–3.0) with at least 80% power with an overall sample size of 50 (assuming *n* = approximately 25 for each sex). We assumed a two-tailed comparison because there was no a priori reason for a given sex to respond better to the active drug. Again, as a supplemental predefined analysis strategy, we assumed that we could strengthen the analyses by removing variance due to nuisance covariates and using longitudinal modeling. Thus, total planned participants needed to increase by an additional 60 subjects, making the total planned recruitment 186. However, because of financial constraints we could only recruit 125 participants. As a result, we determined (prior to any data analysis) that primary effects would be considered suggestive of a positive outcome if a given primary comparison passed the less stringent criterion of *P* < 0.1, rather than *P* < 0.05.

### ***Model Testing***

In order to test the validity of our classifier which stratified participants between high and low risk of pain persistence, we examined the actual responses of the NoTx group at 6 months from entry in the study against the predicted response from our model. To strengthen the validation, in addition to testing the actual response based on our primary outcome measure (phone NRS), we also checked actual versus predicted recovery with other measures of pain intensity.

### ***Identifying Brain Markers for Treatment Effects***

Three regions of interest were defined a priori: the right nucleus accumbens (NAc), medial

prefrontal cortex (mPFC), and right anterior insula (aIns). These were based on previous studies showing that 1. Connectivity strength between mPFC and aIns encodes pain intensity across pain conditions [4], and 2. NAc-mPFC connectivity strength is causally associated with transition from subacute to chronic back pain [5]. Additionally, the NAc has been previously shown to be sex-dimorphic [18]; thus, we reasoned that its connectivity may provide evidence for the sex specificity of treatment response seen for back pain intensity.

Mean time series representative of these regions were extracted from  $10 \times 10 \times 10 \text{ mm}^3$  volumes around previously reported MNI coordinates: NAc (10, 12, -8), mPFC (2, 52, -2), and aIns (42, 14, -6) [4, 5]. The Pearson correlation coefficient ( $R$ ) was computed, three numbers/brain, between BOLD signals from each pair of regions to represent their connectivity strength (extent of information sharing).

#### **Questionnaire-Based Exploratory Endpoints**

Given the large number of additional questionnaires used to assess participants' treatment response, we used dimension reduction methods and network analysis methods to summarize these outcomes. These results are deemed exploratory in nature.

**Missing Questionnaire Data** Missing within-questionnaire items were replaced with the average of the remaining within-questionnaire scores provided that the number of unanswered questions was less than 30% of all items in each scale. If more than 30% of the items were blank, subjects were excluded from statistical tests relevant to the given scale.

**Clustering Analysis** To resolve multicollinearity, dimensionality reduction techniques were applied to pain-related questionnaires at baseline (visit 1). Variable clustering under VARCLUS algorithm (SAS Institute Inc. 2017e) computed Pearson correlation-based similarity between all 19 included measures and assigned these measures to five clusters. VARCLUS is an iterative algorithm that calculates a determination coefficient ( $R^2$ ) between each variable and a cluster (own  $R^2$ ), of

which the variable is a member, and  $R^2$  between each variable and the next most similar cluster (next  $R^2$ ). The  $(1 - R^2)$  ratio is defined as  $(1 - \text{own } R^2)/(1 - \text{next } R^2)$ , where a ratio greater than 1 means that the next closest cluster is more similar than the current cluster.

**Network Analysis** Each of the  $N = 19$  pain-related questionnaire measures was used as a node. On the basis of this division, we constructed an undirected connectivity matrix  $B = R^{N \times N}$  representing all subjects who completed study participation at baseline. First, the Pearson correlation ( $R$ ) coefficient was calculated for each of the possible pairs of nodes to generate a connectivity matrix. Next, this matrix was thresholded to only keep significant connections ( $P < 0.01$ ) and binarized, yielding an undirected adjacency matrix. This process was repeated for each group (LDP + NPX, PLC + NPX, and NoTx) separately at baseline, and at 6 months to investigate long-lasting treatment effects. For network visualization, we used the open-source software Cytoscape.

To characterize the structure of these adjacency matrices, we computed modularity using the Brain Connectivity Toolbox (BCT). A module can be defined as a set of nodes that are densely connected among themselves but sparsely connected to other parts of the network. Modularity quantifies how well defined these densely connected sets of nodes are within the network. From the different modularity algorithms available, we chose to use the fast and accurate multi-iterative generalization of the Louvain method, provided within BCT. Using this technique, we obtained a single unitary value between 0 and 1 representing the modularity of each network, where values closer to 1 indicate highly structured systems and values closer to 0 represent random networks. In order to deal with potential modularity degeneracy, modularity was computed over 100 repetitions and the average of these iterations was used as the final modularity measure.

We further investigated the global network structure by examining changes in connectivity on the non-binarized network from baseline to 6 months and averaging these changes over the entire network to obtain mean  $\Delta R$ .

**Group Differences** Modularity and mean  $\Delta R$  values were compared between groups using a permutation test. First, for each pair-wise measure, the difference between the two groups was calculated as the actual group difference. Second, we identified the lowest common number of subjects and resampled the combined pool of the two conditions into two new groups. The values of these two resampled groups were calculated next. This process was repeated 10,000 times to generate a null distribution of the mean difference between the groups. The  $P$  value of the actual group difference was calculated as the chance probability from the mean in the null distribution.

**Software** Analyses were done using Python 3, R, JMP Pro 13.2 (SAS Institute, Cary, NC), SPSS version 25, and MATLAB 2016a.

**Safety Analysis** Safety analyses included all participants who received at least one dose of the study drug. Adverse events were regularly reported to the data safety monitoring board.

## RESULTS

A total of 125 participants with SBP were screened between February 9, 2015 and March 28, 2018. Of these, 72 were considered eligible for entry into the study (see Supplementary Material Note S2 for inclusion and exclusion criteria) and were stratified on the basis of previously reported neuroimaging biomarkers (see model details in Supplementary Material Note S3) as either having high-risk or low-risk of pain persistence. Sixty-one were identified as at high risk of pain persistence and were randomized in a 1:1 ratio to either LDP + NPX ( $n = 31$ ) or PLC + NPX ( $n = 30$ ). Eleven participants were found to be at low risk of pain persistence, were not given any treatment, and were simply observed for the duration of the study (6 months). As a result of technical failures noted only part way through the study, some participants (18/72) did not fulfill the minimal back pain intensity specified in the

protocol. Additionally, not all participants had documented clinical symptoms of radiculopathy and a small number of participants showed evidence of cannabis use.

Figure 1b describes the trial profile and subject dropouts and distributions across the three study arms. The groups were well balanced with regard to all demographic and clinical variables (age, sex, ethnicity, body mass index, income, education, pain duration; Supplementary Material Table S1).

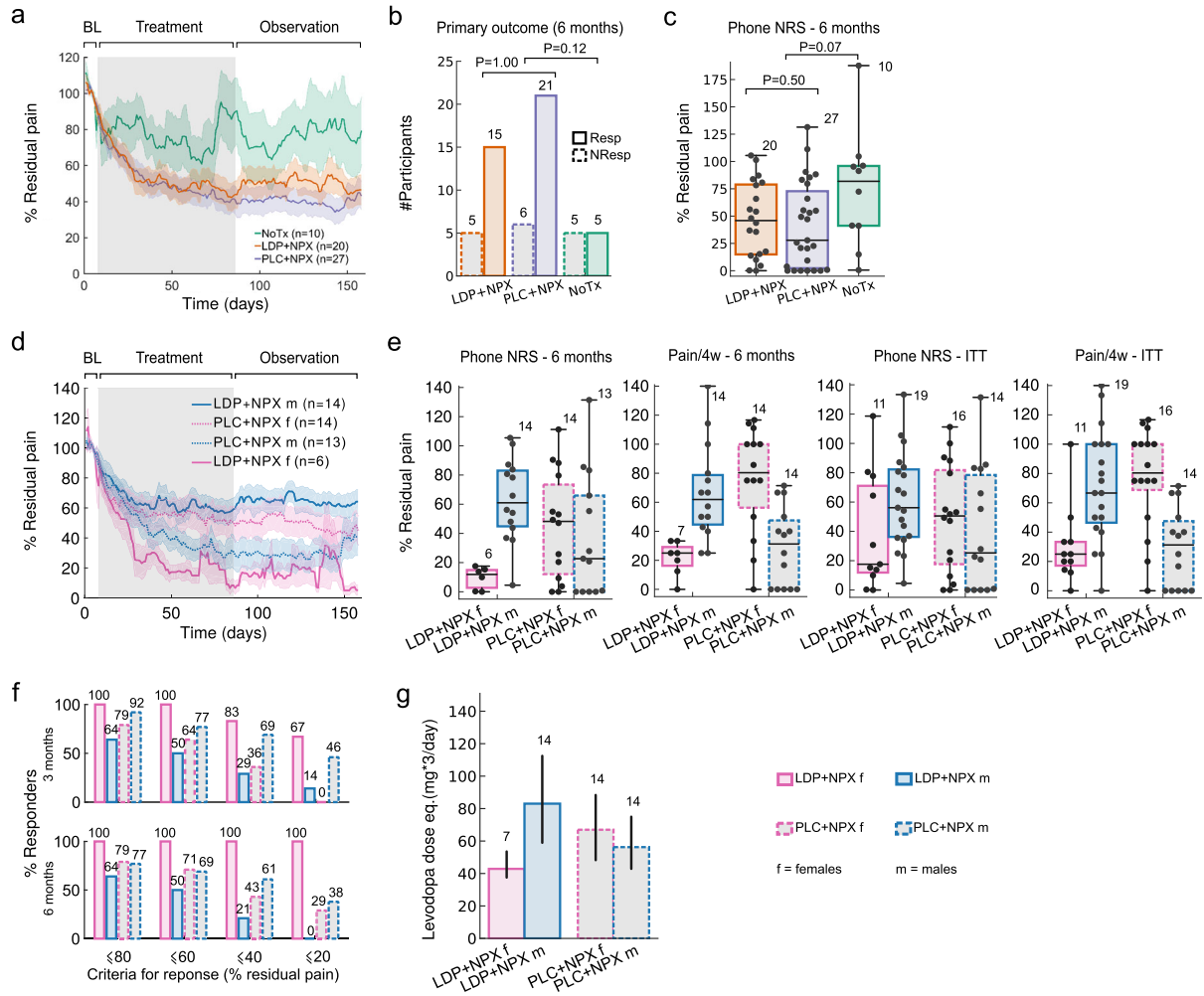
## Efficacy

### *Effect of Treatment on Prevention of Transition to Chronic Pain*

Of the 31 participants randomized to LDP + NPX, 21 completed the full 6 months of the study, but one had missing primary outcome data. Of the remaining 20 participants, 15 (75%) met the criterion for being a responder (pain reduction at study endpoint of at least 20% compared to baseline). In the PLC + NPX group, 28 of the 30 participants completed the full 6 months of the trial, one had missing primary outcome data, and 21 (78%) met this criterion for being a responder. Thus, for both groups, the majority of participants responded to treatment with no significant difference in the rate of transition between groups ( $P = 1.0$ , Fig. 2b). At 3 months, the end of the treatment period, there also were no significant differences in the rate of transition between groups.

### *Effect of Treatment on Pain Intensity*

Back pain intensity, evaluated by means of daily pain reporting on a 0–10 NRS, showed a marked decrease in both treatment groups. At 6 months, the LDP + NPX group had a  $53 \pm 8\%$  (mean  $\pm$  SEM) reduction in back pain intensity; the PLC + NPX group demonstrated a  $58 \pm 7\%$  (mean  $\pm$  SEM) reduction in back pain intensity, with no significant differences between treatment groups ( $P = 0.5$ , Fig. 2c). No significant differences were also observed after 3 months of treatment (Table 1).



◀ **Fig. 2** Back pain time course, response rates, and treatment by sex interaction. **a** Line plot depicts daily average residual back pain trajectories, per study arm: NoTx (green), LDP + NPX (orange), and PLC + NPX (purple) during baseline, treatment and post-treatment observation phases (displayed are 7, 78, and 73 days of ratings for each phase, respectively). There was an interaction between treatment and time (repeated measures mixed model ANOVA,  $F(314,8478) = 1.54$ ,  $P < 0.001$ ), with NoTx residual pain being greater than LDP + NPX ( $P = 0.09$ ) and greater than PLC + NPX ( $P = 0.02$ , post hoc Dunnett test). Gray background indicates treatment phase, shaded areas  $\pm$  SEM. **b** Response rates in the LDP + NPX group (75% of subjects were responders) were not different from PLC + NPX (78% were responders) (Fisher's exact test,  $P = 1.00$ ; with response criterion of greater than 20% reduction in mean pain intensity from baseline to 6 months, mean pain over 1 week of ratings). A trend for superior response rates was seen in the treated population, relative to NoTx (Fisher's exact test,  $P = 0.12$ ). **c** Residual pain at 6 months (phone NRS, 1-week averaged pain ratings relative to baseline) was not statistically different between study arms but showed a trend (Kruskal–Wallis test,  $P = 0.15$ ). Comparing between NoTx and the two treatments together, treated individuals presented larger improvement in pain (Mann–Whitney  $U$  test,  $P = 0.07$ ; medians, quartiles and ranges are shown). **d** Line plot depicts average residual daily pain trajectories for female and male participants separately, per treatment arm. There was an interaction between time, treatment, and sex ( $F(157,6751) = 3.105$ ,  $P < 0.001$ , repeated measures mixed model ANOVA), with LDP + NPX female participants presenting lower residual pain than LDP + NPX male participants ( $P < 0.05$ ), according to post hoc Tukey test. Shaded areas represent  $\pm$  SEM. **e** At 6 months, four measures indicated sex by treatment interaction: 1. residual pain (phone NRS) showed this interaction (two-way ANOVA,  $F(1,43) = 8.16$ ,  $P = 0.007$ ); 2. back pain over 4 weeks (Pain/4w) confirms the finding (two-way ANOVA,  $F(1,45) = 24.59$ ,  $P = 0.00001$ ); 3. Modified intent to treat analysis (ITT), including all subjects with valid data who were randomized in the study, showed a trend for 1-week average phone NRS (two-way ANOVA,  $F(1,56) = 2.33$ ,  $P = 0.13$ ); 4. ITT for 4-week pain showed a robust interaction (Pain/4w, two-way ANOVA,  $F(1,56) = 26.02$ ,  $P = 0.000004$ ). Box plots show median, quartiles, and ranges; while numerals indicate number of subjects. **f** Response rates at both 3 and 6 months across different criteria. LDP + NPX showed complete recovery across all thresholds in all female participants at 6 months, while PLC + NPX was less effective. **g** In LDP + NPX treated individuals, the average maximum administered dose (given three times/day) was lower for female participants than male participants (two-tailed  $t$  test,  $P = 0.07$ ). Just as with pain levels, a treatment by sex interaction was seen in the levels of drug doses received (two-way ANOVA,  $F(1,45) = 4.54$ ,  $P = 0.04$ ). Levodopa dose equivalence was computed according to number of placebo tablets. BL baseline, LDP + NPX levodopa/carbidopa + naproxen, PLC + NPX placebo + naproxen, NoTx no-treatment, Resp responders, NonResp non-responders

### Effect of Sex on Prevention of Transition to Chronic Pain

A determination of the importance of sex in affecting the response to treatment was a pre-specified key secondary endpoint. These data are shown in Fig. 2d–g and Table 2 which demonstrates the responses in the male and female participants of each of the treatment groups. Of the 30 participants in the LDP + NPX group, 11 were female; of the 30 participants in the PLC + NPX group 16 were female. Among the female participants in the LDP + NPX group, seven completed the full 6 months of the study and all six with available data met the criterion of being a responder. Of the male participants, 14 completed the full 6 months of the study and nine met the criterion of being a responder;  $P = 0.26$  for response difference between sexes in the LDP + NPX group, which

was confirmed by sensitivity analysis utilizing BOCF.

### Effect of Sex on Pain Intensity

With regard to the magnitude of reduction in mean pain intensity, ANOVA demonstrated a sex by treatment interaction ( $P = 0.007$ ) from baseline to 6 months. The female participants in the LDP + NPX group reported a  $90 \pm 3\%$  (mean  $\pm$  SEM) reduction in pain compared to  $37 \pm 8\%$  (mean  $\pm$  SEM) in the male participants (PP analyses). Similar comparisons between sexes in the PLC + NPX group showed no significant interaction by sex. The sex by treatment interaction was replicated using both the entire time-series of daily pain intensity and 4-week average pain intensity changes (Fig. 2d, e, respectively). These results were confirmed by

**Table 1** Group average back pain over time

	BL	3 months	6 months
LDP + NPX ( <i>n</i> = 20)			
Mean ± SEM	4.7 ± 0.4	2.3 ± 0.4	2.3 ± 0.4
95% CI	3.9–5.5	1.4–3.2	1.4–3.3
PLC + NPX ( <i>n</i> = 27)			
Mean ± SEM	5.1 ± 0.3	2.4 ± 0.4	2.2 ± 0.4
95% CI	4.4–5.8	1.6–3.2	1.3–3.0
NoTx ( <i>n</i> = 10)			
Mean ± SEM	3.6 ± 0.5	3.2 ± 0.7	2.8 ± 0.7
95% CI	2.4–4.8	1.5–4.8	1.2–4.3

The table summarizes 1-week group-averaged back pain (0–10 scale) at baseline and at 3 and 6 months, as a function of treatment type (mean and standard error, SEM, as well as 95% confidence intervals, CI, are shown). The CIs at 3 or 6 months post-treatment do not overlap with baseline CIs for LDP + NPX and PLC + NPX groups; but they do overlap for the NoTx group  
*BL* baseline, *LDP + NPX* levodopa/carbidopa + naproxen, *PLC + NPX* placebo + naproxen, *NoTx* no-treatment

additional hierarchical linear modeling ( $\beta = -0.4 \pm 0.2$ ;  $t(43) = -2.5$ ;  $P = 0.015$ ) and non-parametric modeling of daily pain (interaction significant at all days after 25 days of treatment; Supplementary Material Fig. S3), as well as in part by modified intent to treat analysis when dealing with missing data.

Potential confounding variables such as baseline pain duration, pain intensity, the spread of pain on the body, and pain-related questionnaires showed no influence on these outcomes (Supplementary Material Fig. S7 and Table S5).

### Exploratory Evaluation of Treatment Effects on Pain-Related Psychological Metrics

Participants' pain characteristics were also assessed by questionnaires. Rather than primarily examining treatment effects on each of

**Table 2** Group average back pain over time by gender

Pain intensity (0–10)	LDP + NPX f ( <i>n</i> = 6)		LDP + NPX m ( <i>n</i> = 14)		PLC + NPX f ( <i>n</i> = 14)		PLC + NPX m ( <i>n</i> = 13)	
	BL	3 months	6 months	BL	3 months	6 months	BL	3 months
Mean ± SEM	4.5 ± 0.9	0.7 ± 0.2	0.4 ± 0.2	4.8 ± 0.4	3.0 ± 0.5	3.2 ± 0.5	5.7 ± 0.4	3.2 ± 0.6
95% CI	2.0–6.9	0.0–1.4	0.0–0.8	3.9–5.7	1.9–4.1	2.1–4.2	4.8–6.5	2.0–4.4

The table summarizes group-averaged back pain at baseline and at 3 and 6 months, as a function of treatment type and sex. Best separation of CIs at 3 or 6 months relative to baseline was seen in female participants (f) treated with LDP + NPX, and in males participants (m) treated with PLC + NPX  
*BL* baseline, *LDP + NPX* levodopa/carbidopa + naproxen, *PLC + NPX* placebo + naproxen, *NoTx* no-treatment



these measures separately (for unitary assessments, see Supplementary Material Table S4), we opted for dimensionality reduction techniques for two reasons. The first is multicollinearity; the second is that most questionnaire scores have only been validated to study chronic pain conditions, and not the transition from subacute to chronic back pain. So, we used a clustering algorithm and constructed a network that summarizes participants' pain-related psychological profile. We then explored treatment effects on this network.

At baseline, the network of pain and psychological metrics was derived from the correlation matrix between all measures across all participants ( $n = 116$ ) (Fig. 3a). Five communities of metrics were identified and labeled according to member scales: 1. pain intensity; 2. pain sensitivity; 3. pain quality; 4. pain affect; and 5. negative affect. Following a threshold to only keep Pearson correlations with  $P < 0.01$ , we observed that these communities were tightly linked with each other at baseline (Fig. 3b). At 6 months, the network structure was disrupted as a function of treatment type (Fig. 3c). Treatment with LDP + NPX decreased correlation strengths throughout the network both in female and male participants (Fig. 3d, f), was associated with increased network modularity (Fig. 4e), and dissociated intensity measures from affective factors (Fig. 3c).

The pain-related psychological network was modified most with LDP + NPX treatment, and specifically in female participants treated with LDP + NPX, with the pain intensity community becoming independent from the rest of the network.

### Objective Brain Functional Correlates for Treatment by Sex Interaction and for Pain Relief

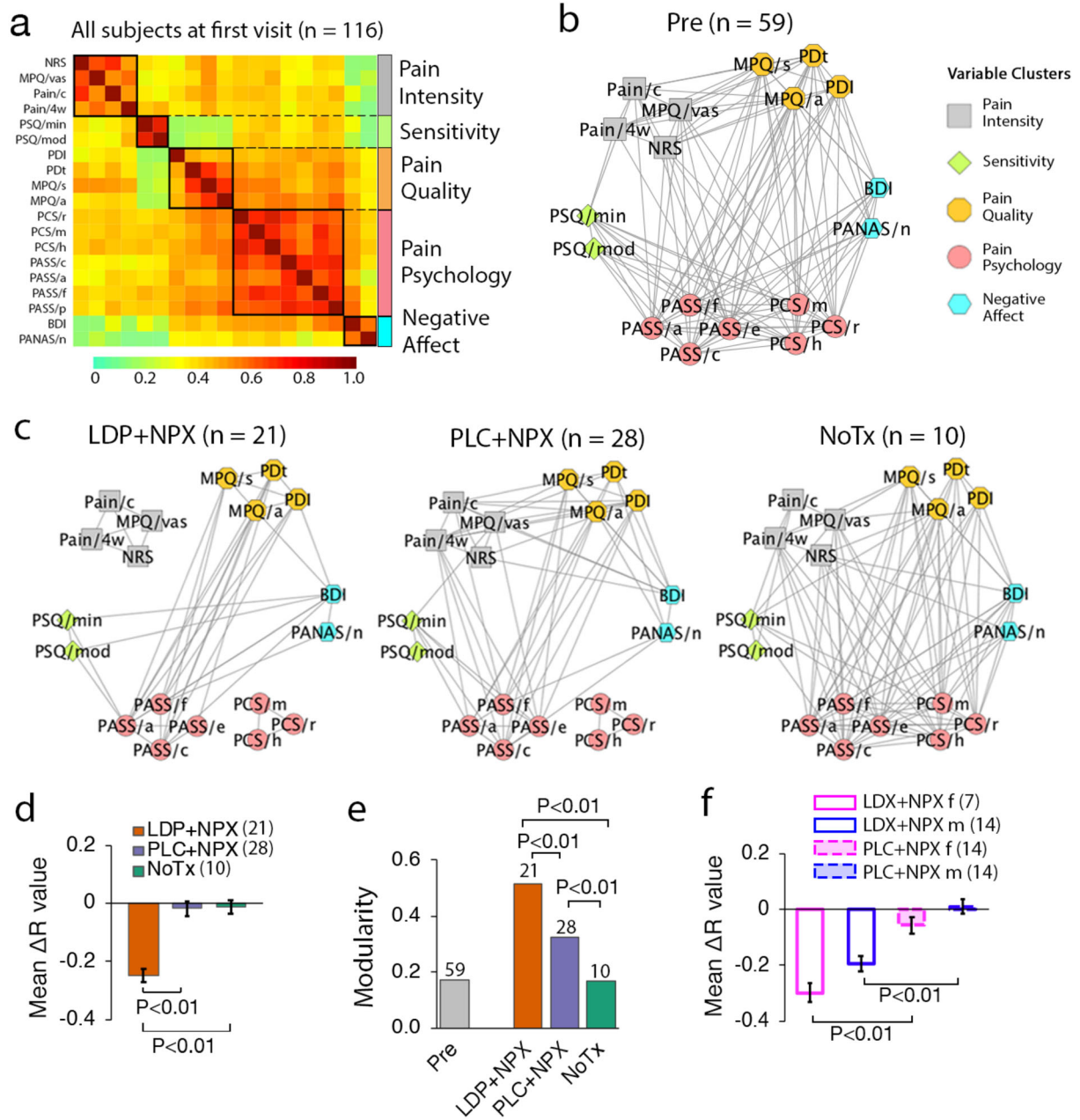
Seeking to test objective neurobiological correlates to the subjective reports of treatment effects, we tested the extent of information sharing between three regions of interest: the right nucleus accumbens (NAc), medial prefrontal cortex (mPFC), and anterior insula (Ins).

We found that these functional connections reflected distinct aspects related to treatment efficacy (Fig. 4a). NAc-mPFC functional connectivity presented sex dependency at baseline ( $P = 0.03$ ) and, in response to treatment, showed an interaction between sex and treatment at 24 weeks after randomization and 12 weeks after treatment cessation (two-way ANOVA,  $P = 0.05$ , Fig. 4a). It should be noted, however, that the post-treatment functional connectivity changes are driven mainly by PLC + NPX and are in opposite directions in male and female participants; thus, these are complex changes that do not directly reflect pain relief. Across study groups, Ins-NAc ( $P = 0.05$ ), but not Ins-mPFC ( $P = 0.16$ ), functional connectivity changes over 6 months were associated with residual pain at 24 weeks (Supplementary Material Fig. S4a, b).

Therefore, predefined neuroimaging-related biomarkers yielded unbiased and objective correlates for treatment effects and sex dependence.

### Relationship Between Brain-Derived Risk of Chronic Pain and Treatment Outcomes

Our brain model predicting risk for chronic pain, calculated at the time of entry into the study, successfully estimated residual pain 6 months later in the NoTx arm based on the main outcome measure (phone NRS,  $P = 0.05$ ) as well as across other pain intensity measures (Fig. 4b, top row). These associations were clearly disrupted in both LDP + NPX and PLC + NPX treatment arms, indicating that both treatments successfully disrupted the predicted natural trajectory of pain (Fig. 4b, middle and bottom rows). Given that our predictive model was validated in the NoTx group, we can approximate the expected recovery rate in the treated subjects from the model. The model predicts that fewer than 25% of the high-risk group (treated with LDP + NPX or PLC + NPX) should have been in the recovering category, while both of these treatments resulted in approximately 75% belonging to the recovering category. Thus, these results validate our



**Fig. 3** Treatment impact on pain-related psychological properties. **a** Correlation matrix for 19 questionnaire subscale measures. Clustering analysis identified 5 communities, based on data from 116 subjects, at the first visit. Communities were labeled according to their component measures: pain intensity, pain sensitivity, pain quality, pain affect, and negative affect. **b** Pre-treatment network graph displays interrelations between the five communities, highlighting within- and inter-cluster associations. Edges represent correlations with  $P < 0.01$ . **c** At 6 months the network structure was disrupted in treated groups, whilst remaining intact in the NoTx group. Specifically, pain intensity and catastrophizing communities dissociated from the network in the LDP + NPX group. **d** LDP + NPX treatment decreased the strength of correlations ( $\Delta R$ ) between measures, relative to PLC + NPX and NoTx groups. **e** Network modularity increased in the treatment groups at 6 months, and was largest in LDP + NPX. **f** Relative to PLC + NPX, LDP + NPX decreased the strength of network correlations in both male and female participants. In **d** and **f**, error measures were based on permutation testing (10,000 repeated random resampling); error bars are SEMs. LDP + NPX levodopa/carbidopa + naproxen, PLC + NPX placebo + naproxen, NoTx no-treatment, NRS numerical rating scale, MPQ McGill Pain Questionnaire—Short Form: visual analog scale (MPQ/vas), sensory (MPQ/s) and affective components of pain (MPQ/a); Pain/c current pain (PainDETECT), Pain/4w (PainDETECT) average pain over the past 4 weeks, PSQ Pain Sensitivity Questionnaire: minor (PSQ/min) and moderate (PSQ/mod), PDI Pain Disability Index, PDt PainDETECT, PCS Pain Catastrophizing Scale: rumination (PCS/r), magnification (PCS/m), and helplessness (PCS/h); PASS Pain Anxiety Symptoms Scale: cognitive suffering (PASS/c), escape-avoidance behaviors (PASS/e), fear of pain (PASS/f), and physiological symptoms of anxiety (PASS/a); BDI Beck Depression Inventory, PANAS Positive and Negative Affect Scale: negative (PANAS/n)

classifier and also provide additional confidence for treatment efficacy.

### Pain Relief is Sustained at 3.3 years

We queried participants at 3.3 years (average) from the trial start. Forty participants could be reached. There was no change in back pain levels between 6 months (end of the trial) to the period 3.3 years after start of trial when evaluated for all participants in the treatment arms, for each subgroup of type of treatment, and for subgroups of sex by type of treatment (Fig. 5 and Supplementary Material Fig. S5). These data extend in time the original observation of blocking transition to chronic pain from 3 months to about 3 years.

We also tested the sex by treatment interaction at 3.3 years from the trial start. This interaction was not sustained (two-way ANOVA,  $P = 0.2$ ), either because of a disruption of this pattern in time or the limited number of subjects we could query (Fig. 5 and Supplementary Material Fig. S5).

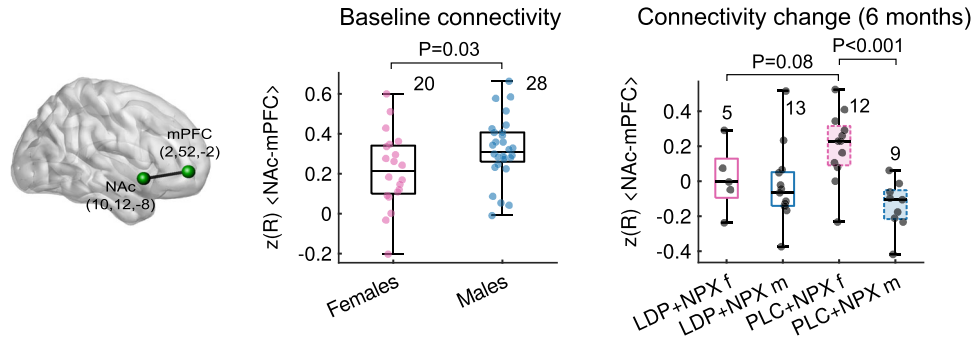
### Safety and Adverse Events

Adverse events were reported by 19 (61.3%) and 17 (56.7%) participants in the LDP + NPX and

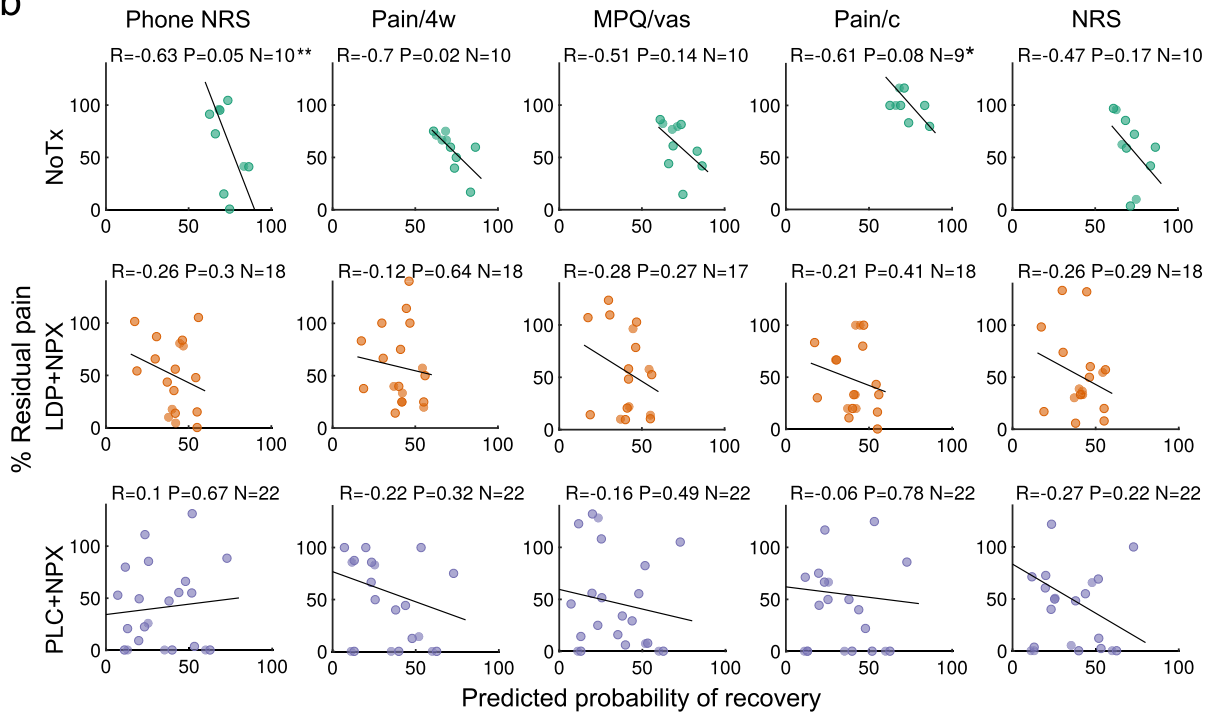
PLC + NPX groups, respectively, with similar incidence of specific adverse events between treatment types (Table 3). Both treatment arms presented adverse events similar to those commonly observed for non-steroidal anti-inflammatory drugs (NSAIDs) [31], and only one case was attributed specifically to LDP. Three LDP + NPX participants reported serious adverse events. Only one of these participants completed the study; he experienced worsening angina and coronary artery disease at the end of treatment phase but chose to continue participation. One male participant was diagnosed with bleeding from esophageal varices during the post-treatment phase and was discontinued from the study. A female participant became pregnant during the treatment period. She stopped treatment, was discontinued from the study, and at a later date had a miscarriage. All serious adverse events were judged to be unrelated to the medication.

Basic hematology and serum chemistry laboratory values at baseline were not different between groups (Supplementary Material Table S8). Heart rate, blood pressure, and body temperature, measured at all visits, were not different between groups and were not influenced by treatments (Supplementary Material Figure S5).

**a**



**b**



◀**Fig. 4** Predefined brain properties reflected sex-dependent treatment effects and successfully predicted future pain. **a** Functional connectivity strength between NAc and mPFC (NAc-mPFC;  $z$ -transformed correlation strength,  $z(R)$ ) was weaker in female participants at baseline (Mann–Whitney  $U$  test,  $P = 0.03$ ). At 6 months, there was an interaction between sex and treatment on the extent of change in NAc-mPFC functional connectivity strength (two-way ANOVA,  $F(1,35) = 5.04$ ,  $P = 0.031$ ). NAc-mPFC functional connectivity change was different between sexes in the PLC + NPX group (post hoc Mann–Whitney  $U$  test,  $P < 0.001$ ). **b** Scatter plots depict model-based predicted probability of recovery versus residual back pain at 6 months, for each treatment type, and for 5 measures of back pain intensity. There is a consistent negative correlation between predicted probability of recovery and residual pain in the NoTx group, across pain intensity measures, indicating that the predictive model used to stratify patients a priori performed well (top panel). Such associations were absent in both LDP + NPX and PLC + NPX treatment groups (lower two panels), suggesting that treatment successfully disrupted what would have been the natural pain trajectories for these participants. \*\*One data point of phone NRS at approximately 190% in the NoTx group is not displayed because of visual constraints, but is included in the regression line; \*One Pain/c outlier (with residual pain of approximately 600%) in the NoTx group was removed; Pearson  $R$  is displayed. LDP + NPX levodopa/carbidopa + naproxen, PLC + NPX placebo + naproxen, NoTx no-treatment, NRS numeric rating scale, MPQ/vas visual analogue scale of McGill questionnaire, Pain/c current pain (PainDETECT), Pain/4w average pain over the past 4 weeks

## DISCUSSION

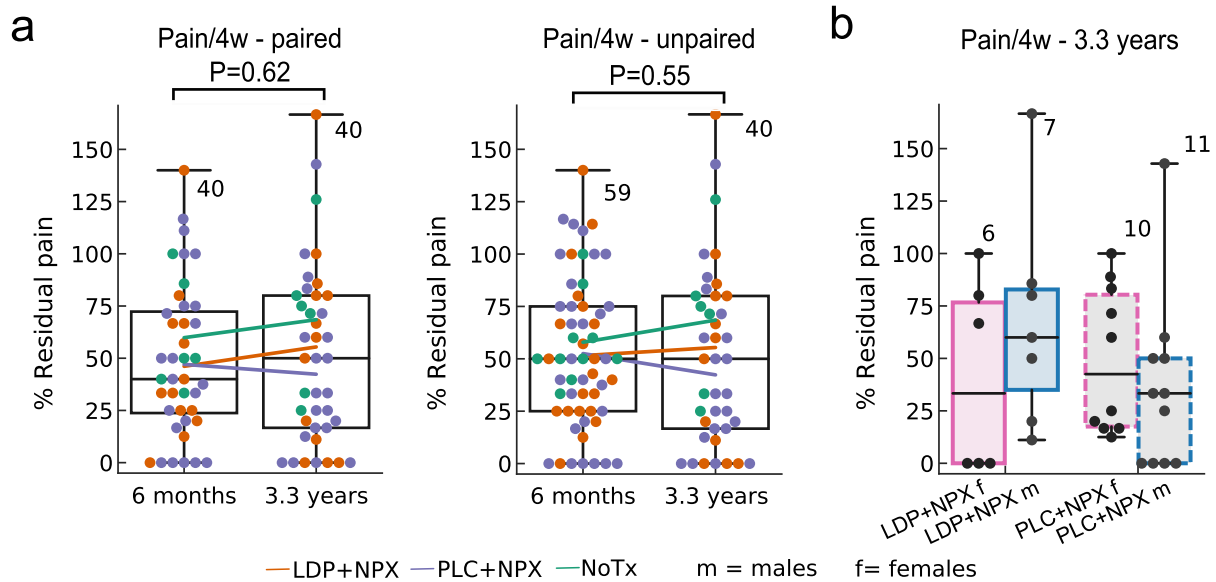
In this proof-of-concept clinical trial in SBP individuals, supported with neuroimaging-based predictors and outcome correlates, we failed to demonstrate the prevention of transition to chronic pain in the group treated with LDP + NPX compared to the group treated with PLC-NPX (primary endpoint) and no significant differences in pain intensity over time were seen between these groups. However, in prespecified analyses focused on evaluating the effect of sex on treatment, we did find a marked difference between female and male participants in response to LDP + NPX, which was highly

significant, with large pain reduction in female participants on this combination treatment. The robustness of this result was supported by all but one of the sensitivity analyses. In addition, we found that the LDP + NPX treatment appeared safe and generally well tolerated. Our exploratory analyses demonstrated that 1. this combination treatment dissociated pain intensity components from the pain-related psychological network; 2. brain functional connectivity reflected sex dimorphism, therapy type, and pain relief, providing complementary objective measures for efficacy; 3. brain parameters determined at the time of entry into the study predicted the risk of back pain 6 months later in the NoTx group but not in the treated subjects, which provides additional independent and objective evidence regarding treatment efficacy.

On the neurobiological level, the treatment specificity seen in the present study between female and male SBP is consistent with the literature regarding MCL being sexually dimorphic [17, 18], which in turn is backed by our female participants having lower resting-state NAc-mPFC functional connectivity strength at entry into the study. Thus, this is the first evidence suggesting sex-specific treatment for back pain, with exceptional outcomes.

In an exploratory analysis we used network theory to profile the pain and related psychological dimensions of individuals with back pain. This analysis provides a global overview of the pain-related psychological profile and shows that these treatments modulate fundamental properties of the network with much less effect on specific questionnaire unitary measures.

As LDP + NPX has not been previously studied in the context of human pain, it was important to carefully document associated adverse events. The incidence and type of observed adverse events were comparable between treatment groups and mainly comprised those commonly related to NSAIDs. Current American and European guidelines for the management of CBP recommend limiting use of NSAIDs because of adverse effects [32]: using them for the shortest duration possible and up to 3 months, respectively. These



**Fig. 5** Back pain relief with treatment was sustained for 3 years post-treatment. **a** Residual pain levels at 6 months were sustained at 3.3 years (paired—Wilcoxon signed-rank test,  $P = 0.62$ , and unpaired—Mann–Whitney  $U$  test,  $P = 0.55$ ) for the entire population, and for each individual subgroup (LDP + NPX,  $P = 0.53$ ; PLC + NPX,  $P = 0.75$ ; NoTx,  $P = 0.25$ , Wilcoxon signed-rank tests). Equivalence testing confirmed similarity of mean

residual pain between 6 months and 3.3 years ( $P = 0.005$ , paired two one-sided  $t$  tests, whole group). **b** At 3.3 years residual pain, based on Pain/4w measure, showed a pattern resembling the sex by treatment interaction observed at 6 months (Fig. 2f), but did not survive statistical testing (two-way ANOVA,  $F(1,30) = 1.74$ ,  $P = 0.20$ )

recommendations are supported by data from a meta-analysis reporting that the mean duration for treating BP with NSAIDs was 5–7 days [33]. In our current study we treated participants for 3 months and found that both PLC + NPX and LDP + NPX were well tolerated with acceptable adverse-event profiles which were similar to those reported for NSAIDs in general [31]. Still, full safety evaluations and long-term effects of these treatments remain to be tested in larger sample sizes. If future studies can replicate our observations, then the benefit-to-risk ratio of 3-month exposure to NSAIDs, with/without LDP, leading to long-term inhibition of the transition to chronic pain may be highly desirable.

### Limitations

The study does have important shortcomings that need highlighting. Foremost, the results remain based on a small number of participants.

Although we planned to enroll 186 participants, only 126 were enrolled. The study population, which consists of people with early-onset back pain, is hard to recruit, because patients tend to seek professional help much later, once the pain is already in its chronic state. Despite the small sample, dense daily monitoring of pain combined with a blinded trial and objective brain biomarkers assessing correlates of primary outcomes make the results unlikely to be a consequence of random effects.

Future studies should ideally also include additional controls, for example: a PLC-only arm, and/or a NoTX arm consisting of participants who are also at high risk for persisting pain. A recent case report [34] shows that in two female participants with back pain and restless leg syndrome, low dose LDP resulted in relief from back pain. Thus, it is also important that future studies compare the efficacy of LDP + NPX in contrast to LDP alone, especially in women. It should be noted that 12% of the participants completing the study did not have

**Table 3** General and specific adverse events

Adverse experience	LDP + NPX, <i>n</i> (%)		PLC + NPX, <i>n</i> (%)	
	(N = 31)		(N = 30)	
General				
Patients with AEs	19	(61.3)	17	(56.7)
Patients with mild AEs	16	(51.6)	16	(53.3)
Patients with moderate AEs	3	(9.7)	1	(3.3)
Patients with severe AEs	1	(3.2)	0	(0.0)
Patients with AEs likely related to treatment	4	(12.9)	3	(10.0)
Patients with AEs possibly related to treatment	12	(38.7)	12	(40.0)
Patients with AEs unlikely related to treatment	5	(16.1)	1	(3.3)
Patients with only one AE	13	(41.9)	13	(43.3)
Patients with multiple AEs	6	(19.4)	4	(13.3)
Total AEs during Tx window	28	(90.3)	21	(70.0)
Total AEs post Tx window	1	(3.2)	1	(3.3)
Specific				
Patients with upper gastrointestinal	6	(19.4)	4	(13.3)
Patients with headache	3	(9.7)	3	(10.0)
Patients with lower gastrointestinal	3	(9.7)	3	(10.0)
Patients with drowsiness	3	(9.7)	1	(3.3)
Patients with lower respiratory infection	2	(6.5)	0	(0)
Patients with influenza	2	(6.5)	0	(0)
Patients with upper respiratory infection	1	(3.2)	3	(10.0)

Incidence of adverse events was small in both treatment groups. In particular, the number of AEs related to medication was similar between LDP + NPX and PLC + NPX, and so was the incidence of specific adverse events. Depicted are specific events that occurred in at least two patients in any one of the treatment arms. Percentages are calculated on the basis of the number of subjects

*LDP + NPX* levodopa/carbidopa + naproxen, *PLC + NPX* placebo + naproxen, *AE* adverse event

a risk estimate of pain persistence (did not undergo brain scans) and were assumed to be at high risk. Although not ideal, we believe this to be a reasonable assumption, given that the expected and observed incidence of high risk was approximately 80%. Still, future studies should also be designed to mitigate such confounds. The study also suffered from multiple

protocol deviations, though these did not appear to impact the final results.

## CONCLUSIONS

The results presented here support the finding that, for blocking tCBP, the combination of

LDP + NPX may be more effective in female participants than NPX itself, a standard approach to CBP management. Additionally, this report provides novel concepts regarding conducting clinical trials, especially for subjective conditions like pain, using personalized brain biomarker-based modeling. These findings need to be replicated in a larger population.

## ACKNOWLEDGEMENTS

We would like to thank all the individuals who participated in this research. We are also thankful to all the lab members who contributed to this study with their time and resources, including Camila B. Pinto and Andrew Vigotsky.

### *Declarations*

**Funding** The study and journal's Rapid Service Fee was funded by the National Institute of Dental and Craniofacial Research (Blue Print Grant number R01DE022746) and in part by National Institute on Drug Abuse (P50 DA044121).

**Authorship** All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

**Author Contributions** DR wrote the manuscript, performed data analyses and collection, and co-led the conduct of the trial. AVA wrote the manuscript, led on study design, supervised conduct of the trial and analyses. TJS wrote the manuscript, led on study design and conduct of the trial, evaluated toxicities and provided clinical oversight. MNB participated in data analyses. PT and MG co-led the conduct of the trial and performed data collection. LH and KW participated in data analyses. BP collected data and participated in data analysis. EVP, TA, RJ, SB, AB collected data and participated in the conduct of the trial. JWG was the trial statistician.

**Disclosures** Diane Reckziegel, Pascal Tétreault, Mariam Ghantous, Kenta Wakaizumi, Bogdan Petre, Lejian Huang, Rami Jabakhanji, Taha Abdullah, Etienne Vachon-Preseau, Sara Berger, Alexis Baria, James W Griffith, Marwan N Baliki, Thomas J Schnitzer and A Vania Apkarian have nothing to disclose.

**Compliance with Ethics Guidelines** Protocol and informed consent forms were approved by Northwestern University IRB as well as National Institute of Dental and Craniofacial Research/National Institutes of Health (NIDCR/NIH). All enrolled participants provided written informed consent. Safety and trial oversight were monitored by an independent data safety monitoring board and a clinical research organization, with NIDCR/NIH oversight. The trial was registered on ClinicalTrials.gov, under registry NCT01951105. This study was conducted in accordance with the Declaration of Helsinki and its further amendments.

**Data Availability** Data from our previous studies are already available on OpenPain ([www.openpain.org](http://www.openpain.org)). Data from the present study will be presented in more than one manuscript and will eventually be made available on OpenPain, upon completion of these manuscripts. The full study protocol is available in Supplementary Material 2.

**Open Access.** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence,



visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

- Murray CJ, Lopez AD. Measuring the global burden of disease. *N Engl J Med*. 2013;369(5):448–57. <https://doi.org/10.1056/NEJMra1201534>.
- Jalal H, Buchanich JM, Roberts MS, Balmert LC, Zhang K, Burke DS. Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. *Science*. 2018. <https://doi.org/10.1126/science.aau1184>.
- Baliki MN, Baria AT, Apkarian AV. The cortical rhythms of chronic back pain. *J Neurosci*. 2011;31(39):13981–90. <https://doi.org/10.1523/JNEUROSCI.1984-11.2011>.
- Baliki MN, Mansour AR, Baria AT, Apkarian AV. Functional reorganization of the default mode network across chronic pain conditions. *PLoS One*. 2014;9(9): e106133. <https://doi.org/10.1371/journal.pone.0106133>.
- Baliki MN, Petre B, Torbey S, et al. Corticostriatal functional connectivity predicts transition to chronic back pain. *Nat Neurosci*. 2012;15(8): 1117–9. <https://doi.org/10.1038/nn.3153>.
- Mano H, Kotecha G, Leibnitz K, et al. Classification and characterisation of brain network changes in chronic back pain: a multicenter study. *Wellcome Open Res*. 2018;3:19. <https://doi.org/10.12688/wellcomeopenres.14069.2>.
- Vachon-Presseau E, Tetreault P, Petre B, et al. Corticolimbic anatomical characteristics predetermine risk for chronic pain. *Brain*. 2016. <https://doi.org/10.1093/brain/aww100>.
- Costa Lda C, Maher CG, McAuley JH, et al. Prognosis for patients with chronic low back pain: inception cohort study. *BMJ*. 2009;339: b3829. <https://doi.org/10.1136/bmj.b3829>.
- Vachon-Presseau E, Centeno MV, Ren W, et al. The emotional brain as a predictor and amplifier of chronic pain. *J Dent Res*. 2016;95(6):605–12. <https://doi.org/10.1177/0022034516638027>.
- Mansour AR, Baliki MN, Huang L, et al. Brain white matter structural properties predict transition to chronic pain. *Pain*. 2013;154(10):2160–8. <https://doi.org/10.1016/j.pain.2013.06.044>.
- Lee M, Manders TR, Eberle SE, et al. Activation of corticostriatal circuitry relieves chronic neuropathic pain. *J Neurosci*. 2015;35(13):5247–59. <https://doi.org/10.1523/JNEUROSCI.3494-14.2015>.
- Ren W, Centeno MV, Berger S, et al. The indirect pathway of the nucleus accumbens shell amplifies neuropathic pain. *Nat Neurosci*. 2016;19(2):220–2. <https://doi.org/10.1038/nn.4199>.
- Chang PC, Pollema-Mays SL, Centeno MV, et al. Role of nucleus accumbens in neuropathic pain: linked multi-scale evidence in the rat transitioning to neuropathic pain. *Pain*. 2014;155(6):1128–39. <https://doi.org/10.1016/j.pain.2014.02.019>.
- Fillingim RB, Doleys DM, Edwards RR, Lowery D. Clinical characteristics of chronic back pain as a function of gender and oral opioid use. *Spine*. 2003;28(2):143–50.
- Meucci RD, Fassa AG, Faria NM. Prevalence of chronic low back pain: systematic review. *Rev Saude Publica*. 2015. <https://doi.org/10.1590/S0034-8910.2015049005874>.
- Mogil JS. Sex differences in pain and pain inhibition: multiple explanations of a controversial phenomenon. *Nat Rev Neurosci*. 2012;13(12):859–66. <https://doi.org/10.1038/nrn3360>.
- Koscik T, Bechara A, Tranel D. Sex-related functional asymmetry in the limbic brain. *Neuropsychopharmacology*. 2010;35(1):340–1. <https://doi.org/10.1038/npp.2009.122>.
- Munro CA, McCaul ME, Wong DF, et al. Sex differences in striatal dopamine release in healthy adults. *Biol Psychiatry*. 2006;59(10):966–74. <https://doi.org/10.1016/j.biopsych.2006.01.008>.
- Farrar JT, Young JP Jr, LaMoreaux L, Werth JL, Poole MR. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. 2001;94(2):149–58. [https://doi.org/10.1016/S0304-3959\(01\)00349-9](https://doi.org/10.1016/S0304-3959(01)00349-9).
- Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. *J Pers Assess*. 1996;67(3):588–97. [https://doi.org/10.1207/s15327752jpa6703\\_13](https://doi.org/10.1207/s15327752jpa6703_13).
- Ruscheweyh R, Verneuer B, Dany K, et al. Validation of the pain sensitivity questionnaire in chronic pain patients. *Pain*. 2012;153(6):1210–8. <https://doi.org/10.1016/j.pain.2012.02.025>.
- Tait RC, Pollard CA, Margolis RB, Duckro PN, Krause SJ. The Pain Disability Index: psychometric and validity data. *Arch Phys Med Rehabil*. 1987;68(7):438–41.
- Freyenhagen R, Baron R, Gockel U, Tolle TR. pain-DETECT: a new screening questionnaire to identify

- neuropathic components in patients with back pain. *Curr Med Res Opin.* 2006;22(10):1911–20. <https://doi.org/10.1185/030079906X132488>.
24. Melzack R. The short-form McGill Pain Questionnaire. *Pain.* 1987;30(2):191–7.
  25. Sullivan MJLBPJ. The Pain Catastrophizing Scale: development and validation. *Psychol Assess.* 1995;7:524–32.
  26. McCracken LM, Zayfert C, Gross RT. The Pain Anxiety Symptoms Scale: development and validation of a scale to measure fear of pain. *Pain.* 1992;50(1):67–73.
  27. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol.* 1988;54(6):1063–70. <https://doi.org/10.1037//0022-3514.54.6.1063>.
  28. Baria AT, Mansour A, Huang L, et al. Linking human brain local activity fluctuations to structural and functional network architectures. *Neuroimage.* 2013;73:144–55. <https://doi.org/10.1016/j.neuroimage.2013.01.072>.
  29. Smith SM, De Stefano N, Jenkinson M, Matthews PM. Normalized accurate measurement of longitudinal brain change. *J Comput Assist Tomogr.* 2001;25:466–75.
  30. Ivosev G, Burton L, Bonner R. Dimensionality reduction and visualization in principal component analysis. *Anal Chem.* 2008;80(13):4933–44. <https://doi.org/10.1021/ac800110w>.
  31. Bhala N, Emberson J, Merhi A, et al. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials. *Lancet.* 2013;382(9894):769–79. [https://doi.org/10.1016/S0140-6736\(13\)60900-9](https://doi.org/10.1016/S0140-6736(13)60900-9).
  32. Enthoven WTM, Roelofs PD, Koes BW. NSAIDs for chronic low back pain. *JAMA.* 2017;317(22):2327–8. <https://doi.org/10.1001/jama.2017.4571>.
  33. Machado GC, Maher CG, Ferreira PH, Day RO, Pinheiro MB, Ferreira ML. Non-steroidal anti-inflammatory drugs for spinal pain: a systematic review and meta-analysis. *Ann Rheum Dis.* 2017;76(7):1269–78. <https://doi.org/10.1136/annrheumdis-2016-210597>.
  34. Zeng ZF, et al. Chronic back pain cured by low-dose levodopa: is it a variant of restless legs syndrome? *J Pain Res.* 2018;11:277–9. <https://doi.org/10.2147/JPR.S156894>