

1       **Proteome-wide and mendelian randomisation analyses of chronic widespread pain:**  
2                               **diagnosis, prognosis, and drug target discovery**

3  
4   Li Chen <sup>1,2</sup>, Eoin Kelleher<sup>3</sup>, Ruogu Meng <sup>4</sup>, Duanke Liu <sup>5</sup>, Yuchen Guo <sup>1</sup>, Yunhe Wang <sup>6</sup>,  
5   Yaoqing Gao <sup>6</sup>, Zhe Huang <sup>6</sup>, Zhu Liang <sup>7</sup>, Shuai Yuan <sup>8,9</sup>, Chao Zeng <sup>10</sup>, Guanghua Lei <sup>10</sup>,  
6   Jun Ma <sup>2</sup>, Yanhui Dong <sup>2&</sup>, Anushka Irani <sup>1,11&</sup>, Junqing Xie <sup>1,12\*</sup>, Daniel Prieto-Alhambra  
7   <sup>1,13&</sup>

8   1. Centre for Statistics in Medicine and NIHR Biomedical Research Centre Oxford,  
9       NDORMS, University of Oxford, Oxford, UK

10   2. Institute of Child and Adolescent Health, School of Public Health, Peking University

11   3. Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

12   4. National Institute of Health Data Science, Peking University, Beijing, China

13   5. Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren  
14       Hospital, Capital Medical University, Beijing Key Laboratory of Ophthalmology and  
15       Visual Sciences, Beijing, China

16   6. Nuffield Department of Population Health, University of Oxford, Oxford, UK

17   7. Nuffield Department of Medicine, University of Oxford, Target Discovery Institute,  
18       Center for Medicines Discovery, Oxford, United Kingdom

19   8. Department of Surgery, Perelman School of Medicine at the University of Pennsylvania,  
20       PA, USA

21   9. Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

22   10. Department of Orthopaedics, Xiangya Hospital, Central South University, Changsha,  
23       China

24   11. Division of Rheumatology, Mayo Clinic Florida, Florida, USA

25   12. The Queen's College, University of Oxford, Oxford, UK

26   13. Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, the  
27       Netherlands

28  
29  
30   & Joint senior authorship

31   \*Correspondence to:

32   DPhil. Junqing Xie

33   Email: Junqing.xie@ndorms.ox.ac.uk

- 34 Centre for Statistics in Medicine and NIHR Biomedical Research Centre Oxford, NDORMS,  
35 University of Oxford, Oxford, UK;

36 **Abstract**

37 Chronic widespread pain (CWP) remains challenging due to its heterogenous causes and  
38 complex mechanisms. We analysed 2,923 plasma proteins from 29,254 UK Biobank  
39 participants. We first identified 811 proteins correlated with the presence of CWP cross-  
40 sectionally. We then created a sparse (top 10 proteins) and intricate (all significant proteins)  
41 proteomic-based score (ProtS) for CWP, both outperforming and improving the traditional  
42 clinical score (AUC: 0.82 and 0.88 vs 0.81 individually, and 0.87 and 0.92 in combination).  
43 Prospectively, the ProtS was associated with increased risks of a spectrum of pain traits,  
44 including dimensions from pain onset, progression and intensity, up to 13-years of follow-up;  
45 More importantly, we identified distinct proteomic signatures for nociplastic pain compared  
46 to nociceptive and neuropathic pain. For the individual proteins, carbonic anhydrase 14  
47 (CA14) and leptin appeared as promising casual CWP biomarkers as triangulated by  
48 Mendelian randomisation and colocalization analyses. Lastly, our drug-repurposing analysis  
49 identified ten potential candidates for CWP therapies, calling for further research including  
50 randomised controlled trials. Although no CA14 agonists are currently available, CA14  
51 remains a promising target, warranting further efforts to explore its role in pain modulation.

52 **Introduction**

53 Chronic pain affects 27.5% of the global population and is one of the leading causes of  
54 medical consultation and disability<sup>1,2</sup>, particularly among adults over 65, where joint, back,  
55 and neck pain are common<sup>3,4</sup>. The burden of chronic pain extends beyond physical suffering,  
56 as it poses significant societal costs through lost productivity and increased healthcare  
57 resource utilisation, presenting a major global health challenge<sup>5,6</sup>.

58

59 The International Association for the Study of Pain (IASP) defines pain as an unpleasant  
60 sensory and emotional experience. While acute pain serves a protective function by signaling  
61 potential or actual tissue damage, chronic pain is defined as pain persisting for more than  
62 three months, continuing beyond the normal healing period. Unlike acute pain, which is  
63 adaptive, chronic pain is often maladaptive as it no longer serves a protective role. Indeed,  
64 chronic pain is increasingly recognised as a distinct condition, with the 2019 ICD-11  
65 classification introducing ‘chronic primary pain’, where pain is the principal problem rather  
66 than a result of another disease<sup>7-10</sup>. Chronic pain was traditionally classified as either  
67 neuropathic pain, which originates from a lesion or disease of the somatosensory nervous  
68 system, or nociceptive pain, which results from actual or threatened non-neural tissue  
69 damage<sup>8,9</sup>. However, in 2016, a third category—nociplastic pain—was introduced to describe  
70 pain arising from altered nociception without clear evidence of tissue or nerve damage.  
71 Nociplastic pain, characterised by its widespread and diffuse nature, is particularly  
72 debilitating and is best typified by conditions such as fibromyalgia<sup>11</sup>.

73

74 Despite advances in understanding these pain mechanisms, clinical tools to accurately  
75 diagnose and distinguish between them remain lacking. Existing clinical models do not  
76 account for the molecular underpinnings that could differentiate nociceptive, neuropathic, and  
77 nociplastic pain. Here, proteomics – the large-scale study of proteins – offers a promising  
78 avenue. By examining proteins as biomarkers, proteomics offers promising opportunities to  
79 improve diagnostic accuracy and support personalised therapeutic strategies for chronic pain.

80

81 The field of proteomics has significantly advanced our understanding of molecular  
82 mechanisms in complex diseases like cardiovascular disease, diabetes, and cancer by  
83 integrating genetic, lifestyle, and environmental influences<sup>12</sup>. Given the recognition of pain  
84 vulnerability—where genetic and biological factors predispose certain individuals to  
85 developing chronic pain<sup>13-15</sup>—proteomic analyses could help identify those at higher risk,

86 facilitating early recognition and intervention. Prior biomarker studies in chronic pain have  
87 been limited by small sample sizes and a narrow focus on specific proteins, failing to capture  
88 the full complexity of pain mechanisms<sup>16</sup>. By leveraging a comprehensive proteomic dataset  
89 in a large population-based study, our research addresses this gap and provides an opportunity  
90 to identify novel biomarkers associated with pain vulnerability and distinct pain mechanisms  
91<sup>17</sup>. These innovative approaches may offer a deeper understanding of chronic pain  
92 mechanisms, opening pathways to improved diagnostic assays and novel therapeutic targets.

93

94 We therefore aimed to: (1) evaluate the proteomic signatures distinguishing acute and chronic  
95 pain, identifying unique molecular patterns; (2) assess the diagnostic and prognostic validity  
96 of models based on 2,923 unique plasma proteins in predicting widespread chronic pain,  
97 comparing their performance to conventional clinical risk models, and to explore the added  
98 value of integrating plasma proteins into clinical risk models; and (3) to assess the causal  
99 relevance of key identified proteins and their potential as therapeutic targets for the treatment  
100 of chronic pain.

## 101 **Results**

### 102 **Baseline Characteristics and Pain Phenotypes**

103 We analysed 29,254 participants with plasma proteomics data (Olink) from the UK Biobank,  
104 of whom 44.6% were male, with a median age of 55.8 years (**Supplementary Table 2**). The  
105 comparison of participants with and without proteins is provided in **Supplementary Table 1**.  
106 In the present study, 52.3% of participants reported suffering chronic pain at baseline.  
107 Regarding pain types, 85.2% reported musculoskeletal pain, 22.5% pain in the head and face,  
108 10.7% abdominal pain, and 3.0% experienced widespread pain. Overall, 24.4% of  
109 participants reported at least two pain sites affected. A Venn diagram depicting the overlap  
110 between different chronic pain types, including widespread chronic pain, is provided in  
111 **Extended Data Fig. 1**. In 2019, data on pain mechanisms were collected from a total of  
112 8,225 participants. Among them, 571 reported nociceptive pain, 75 reported neuropathic pain,  
113 4,153 experienced nociplastic pain, and 717 had a diagnosis of fibromyalgia  
114 (**Supplementary Table 2**).

115

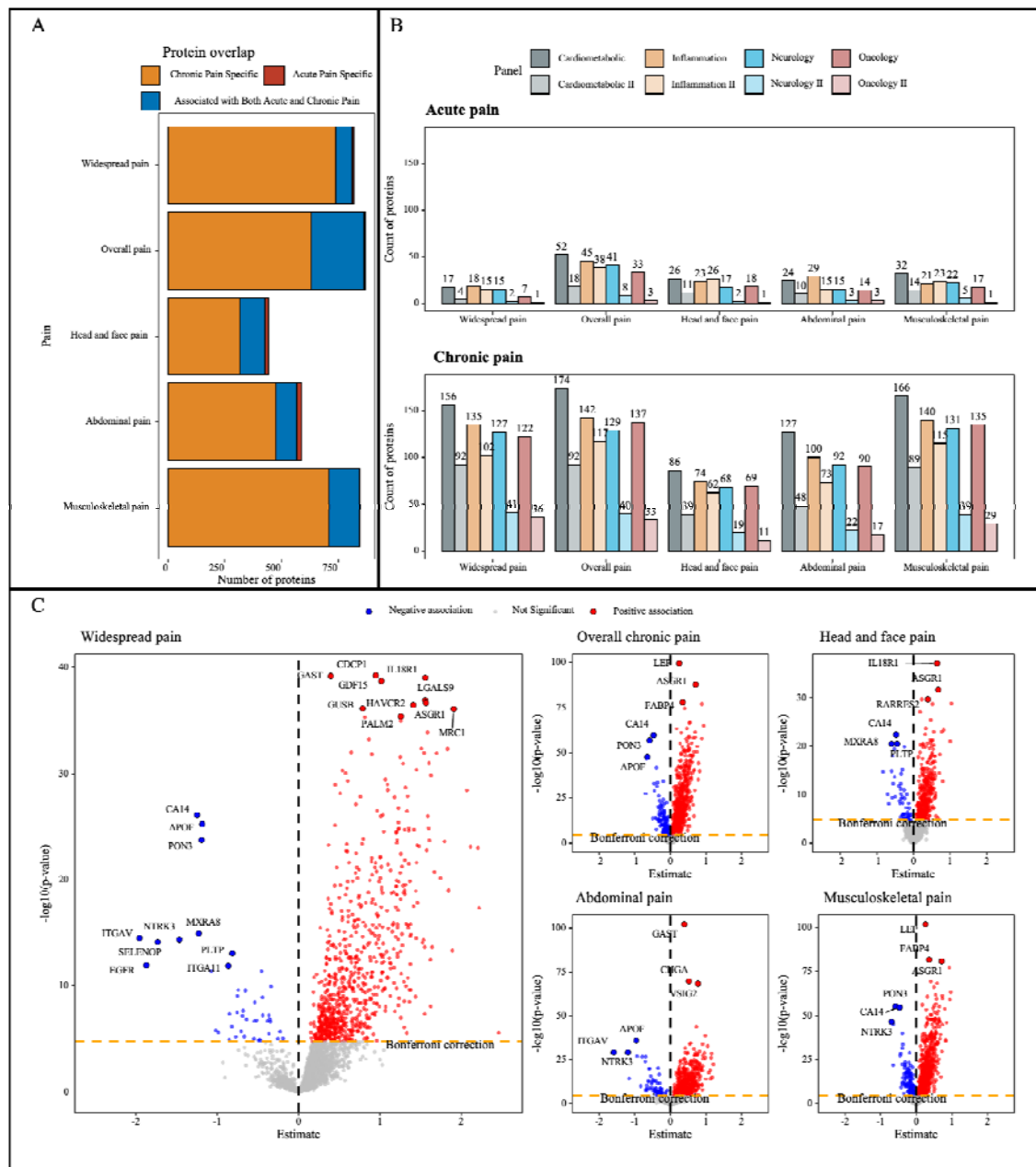
### 116 **Proteomic Associations with Chronic and Acute Pain**

117 We conducted a proteome-wide association study (PWAS) to evaluate the ability of protein  
118 signatures to distinguish chronic pain by comparing protein expression profiles between  
119 individuals with chronic pain and pain-free controls. People with chronic pain showed  
120 widespread and substantial changes in protein expression compared to pain-free controls. In  
121 contrast, those with acute pain showed minimal associations with protein levels (**Fig.1A and**  
122 **Extended Data Fig. 2**). To assess the difference in protein signatures associated with chronic  
123 and acute pain, we analysed the overlap between proteins linked to each condition. Only 27%  
124 (234/864) of the proteins associated with chronic pain overlapped with those linked to acute  
125 pain, underscoring distinct molecular pathways (**Fig.1A and B and Supplementary Table**  
126 **3**).

127

128 In the analysis of widespread chronic pain, the proteins most significantly upregulated  
129 compared to controls were CDCP1, IL18R1, GAST, GDF15, LGALS9, GUSB, ASGR1,  
130 MRC1, PALM2, and HAVCR2. In contrast, the ten most downregulated proteins were CA14,  
131 APOF, PON3, NTRK3, MXRA8, ITGAV, SELENOP, PLTP, EGFR, and ITGA11 (**Fig.1B**). A  
132 total of 811 proteins out of 2,923 (27.7%) exhibited significant associations with widespread  
133 chronic pain. In the PWAS, 864 out of 2,923 proteins (29.6%) exhibited significant  
134 associations with overall chronic pain. Musculoskeletal pain had the highest number of

135 significant associations, with 844 proteins (28.9%), followed by abdominal pain with 569  
 136 proteins (19.5%), and head and face pain with 428 proteins (14.6%) (**Fig. 1B and**  
 137 **Supplementary Table 4**). Several proteins, including APOF, ASGR1, CA14, CA6, MXRA8,  
 138 PLTP, PON3, and RARRES2, were shared across multiple pain types (**Fig. 1C**).



139

140 **Figure 1. Proteomic signature of different types of chronic pain and acute pain.**

141 Notes: A, Comparison of proteins significantly associated with both chronic and acute pain. B, Count  
 142 of proteins significantly associated with specific types of acute and chronic pain, categorised by panel.  
 143 C, Volcano plots show protein associations with different pain types after applying Bonferroni  
 144 correction. Proteins with p-values below 0.05/2923 are highlighted in red (positive) and blue

145 (negative), while nonsignificant associations are represented in gray. The top 10 positively and  
146 negatively associated proteins with the lowest p-values for each pain type are labelled.

147

### 148 **Intricate and sparse proteomic scores (I-ProtS and S-ProtS)**

149 Proteins identified from PWAS for widespread chronic pain and specific chronic pain types  
150 were used to derive intricate proteomic scores (I-ProtS) for chronic pain. Additionally, sparse  
151 proteomic scores (S-ProtS) were generated using the top 10 proteins associated with  
152 widespread chronic pain and specific chronic pain types to evaluate the potential of these  
153 sparse plasma protein signatures in distinguishing specific chronic pain types.

154

155 Higher baseline percentiles of both I-ProtS and S-ProtS correlated with increased prevalence  
156 of chronic pain at baseline, both across overall as well as for specific chronic pain types. The  
157 associations between I-ProtS and overall chronic pain are illustrated in **Extended Data Fig. 3**,  
158 showing a strong correlation coefficient [95% CI]: 0.99 [0.99–1.00]. Across all chronic pain  
159 types, S-ProtS showed a distribution similar to I-ProtS and comparable correlations with  
160 observed event rates. The association between I-ProtS or S-ProtS and the observed rates of  
161 widespread pain differed from that of other specific pain types (**Extended Data Fig. 3**).

162

### 163 **Protein signatures improved classification of widespread chronic pain**

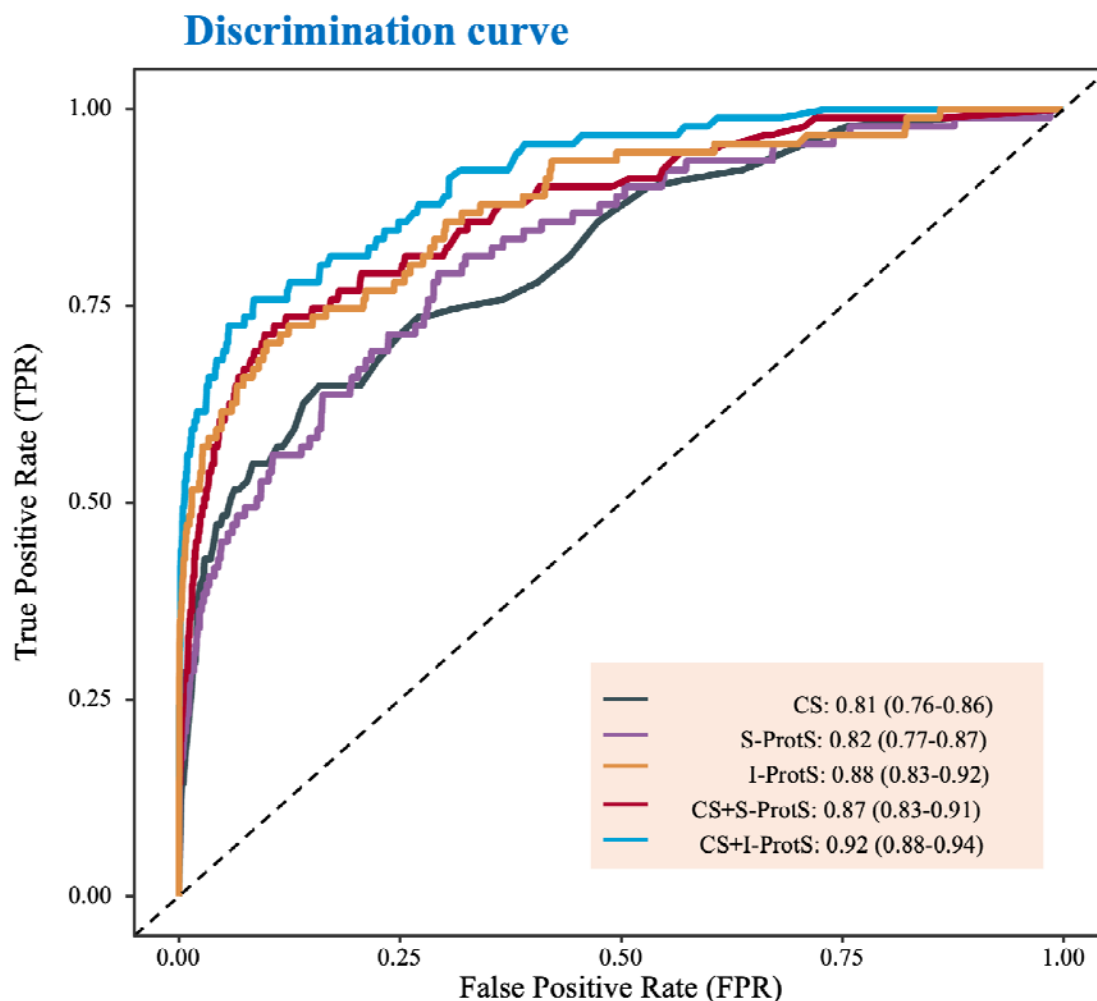
164 Clinical risk models (CS) incorporating previously<sup>13</sup> identified key risk factors like sleep  
165 difficulties, neuroticism, powerlessness, mental health issues, life stressors, and obesity  
166 achieved a median Area Under the Curve (AUC) of 0.81 (95% CI: 0.76–0.86) for widespread  
167 chronic pain. A protein-only model based on S-ProtS showed comparable performance, with  
168 an AUC of 0.82 (95% CI: 0.77–0.87). The I-ProtS model achieved numerically higher  
169 performance, with an AUC of 0.88 (95% CI: 0.83–0.92). Incorporating I-ProtS into the CS  
170 model enhanced discrimination, yielding an AUC of 0.87 (95% CI: 0.83–0.91) for  
171 widespread chronic pain. The combined I-ProtS and CS model reached the highest  
172 discrimination, with an AUC of 0.92 (95% CI: 0.88–0.94) (**Fig. 2A**). To assess the  
173 interpretability of the classification models, we evaluated their calibration and performed  
174 decision curve analyses. For widespread chronic pain, the combined models demonstrated  
175 good calibration, as evidenced by a calibration slope of 0.97 (95% CI 0.85–1.10), indicating  
176 strong alignment between predicted and observed risks (**Extended Data Fig. 4**).

177



178 CS models for chronic pain had AUC ranging from 0.66 to 0.76, with the highest  
179 performance observed for abdominal pain (AUC = 0.76, 95% CI: 0.73–0.79). Across the  
180 three pain types, the I-ProtS demonstrated comparable performance to the CS model.  
181 However, the sparse proteomic score (S-ProtS) model did not achieve superior or comparable  
182 performance relative to the CS model for any chronic pain types (**Extended Data Fig. 4**).  
183 Incorporating S-ProtS into the CS model significantly enhanced discrimination for all chronic  
184 pain types, with an AUC of 0.69 (95% CI: 0.67–0.70) for Musculoskeletal pain and the best  
185 performance for abdominal pain (AUC = 0.81, 95% CI: 0.78–0.83). The addition of I-ProtS  
186 to the CS model further improved classification performance beyond that of S-ProtS, with an  
187 AUC of 0.91 (95% CI: 0.90–0.92) for head and face pain (**Extended Data Fig. 4**). The I-  
188 ProtS-only model showed a net benefit comparable to that of the CS-based models, while the  
189 combined models (incorporating I-ProtS or S-ProtS) provided a greater net benefit than  
190 models based solely on either CS or I-ProtS alone (**Extended Data Fig. 4**).

191



192

193 **Figure 2. Diagnostic performance of proteomic score, clinical score, and combined**

194 **Notes:** this figure presents the ROC curves and Area Under the Curve (AUC) for widespread  
195 chronic pain. Higher AUC values indicate improved model performance in distinguishing  
196 individuals with widespread chronic pain from those without. **Abbreviations:** CS: clinical  
197 risk model; S-ProtS: sparse proteomic score; I-ProtS: intricate proteomic score; CS+S-ProtS:  
198 combined model with clinical model and S-ProtS; CS+I-ProtS: combined model with clinical  
199 model and I-ProtS.

200

### 201 **Prospective Prediction of Pain Progression and Onset**

202 The stability and individual changes in pain site numbers between baseline and the two  
203 follow-up visits were shown in **Extended Data Fig. 5**. We assessed the predictive power of I-  
204 ProtS and S-ProtS for predicting new widespread pain, pain mechanisms, and pain

205 progression. Both I-ProtS and S-ProtS demonstrated good predictive discrimination for  
206 fibromyalgia, with AUCs of 0.76 and 0.74, respectively (**Fig. 4A**). The ProtS showed better  
207 calibration for the nociplastic pain and fibromyalgia (**Supplementary Fig.1**).

208

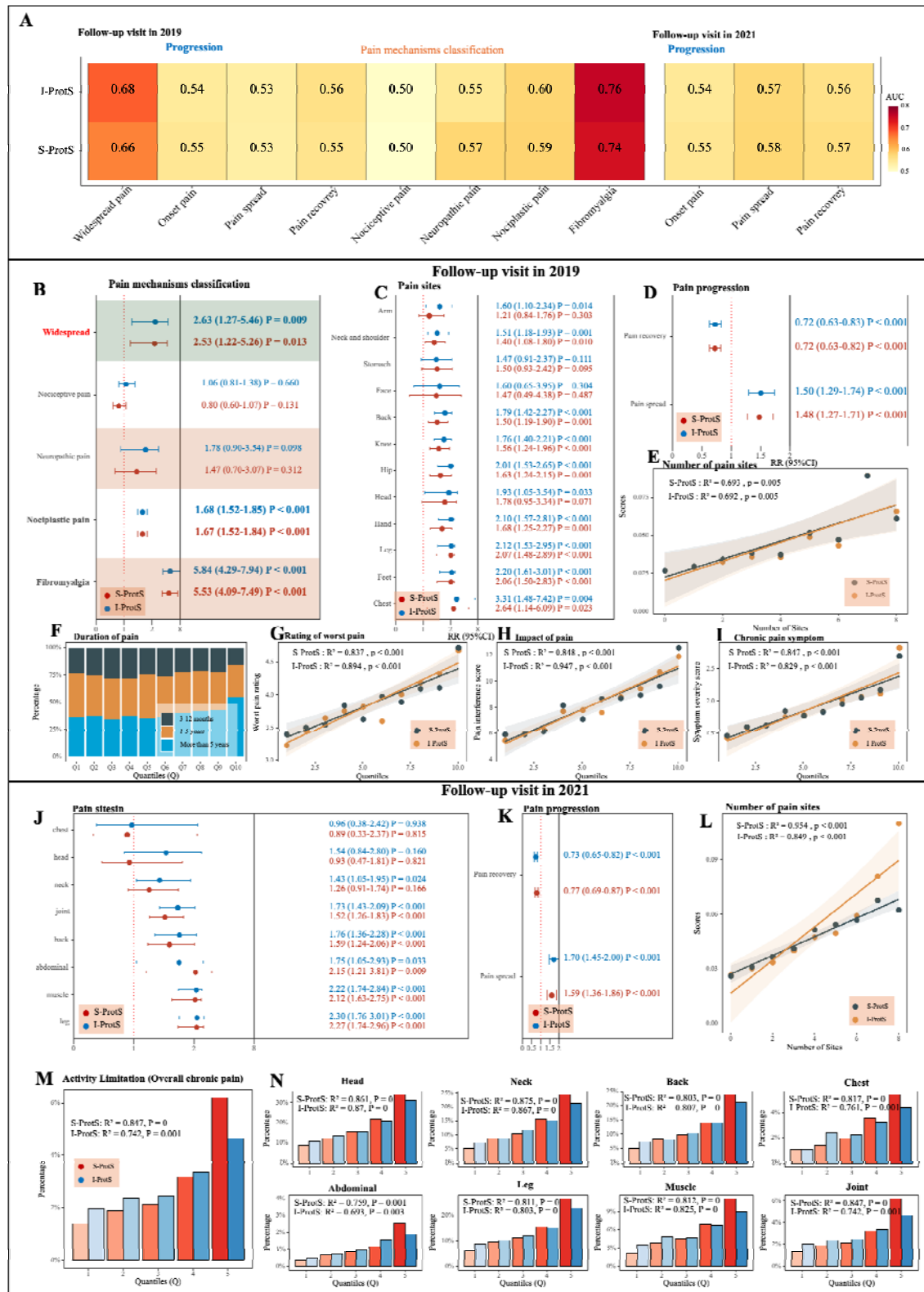
209 Both I-ProtS and S-ProtS for widespread chronic pain were significantly associated with the  
210 onset of widespread chronic pain at follow-up visits (**Fig. 4B**). In the pain mechanism  
211 classification, I-ProtS and S-ProtS were significantly associated with nociplastic pain and  
212 fibromyalgia. However, for nociceptive pain and neuropathic pain, no significant associations  
213 were observed for either I-ProtS or S-ProtS (**Fig. 4B**).

214

215 For the onset of specific pain sites, I-ProtS and S-ProtS also showed significant positive  
216 associations, although these were weaker for certain pain sites, such as headache, facial pain,  
217 and abdominal pain, compared to other chronic pain sites (**Fig. 4C and J**). Both follow-up  
218 visits demonstrated a positive correlation between proteomic scores and pain spread, and a  
219 negative correlation with pain recovery. Higher proteomic scores were associated with an  
220 increased number of pain sites (multisite chronic pain), while lower scores corresponded with  
221 fewer pain sites and greater pain recovery (**Fig. 4D, E, K, and L**). The proteomic scores, both  
222 I-ProtS and S-ProtS, demonstrated a strong association with the number of pain sites at the  
223 2019 follow-up (I-ProtS:  $R^2=0.692$ ,  $P=0.005$ ; S-ProtS:  $R^2=0.693$ ,  $P=0.005$ ) and an even  
224 stronger association at the 2022 follow-up (I-ProtS:  $R^2=0.849$ ,  $P<0.001$ ; S-ProtS:  $R^2=0.954$ ,  
225  $P<0.001$ ) (**Fig. 4C and G**).

226

227 Proteomic scores exhibited a monotonic increase with pain duration, rating of worst pain,  
228 pain impact, and pain symptom severity nine years later (**Fig. 4F-I**). The percentage of  
229 participants reporting activity limitations due to chronic pain increased across higher  
230 proteomic score quantiles (I-ProtS and S-ProtS) (**Fig. 5L**). S-ProtS showed comparable  
231 associations with activity limitations related to chronic pain sites, similar to those observed  
232 with I-ProtS (**Fig. 4N**).



233

234 **Fig. 4: Prospective associations of pain-specific proteomic scores with mechanism pain**  
235 **classification, pain onset, progression, multisite chronic pain, duration, intensity, impact,**  
236 **and activity limitations in longitudinal data.**

237 Notes: The 2019 follow-up refers to the online follow-up conducted in 2019, and the 2022  
238 follow-up refers to the one conducted in 2022. Panel A shows the AUC of I-ProtS and S-  
239 ProtS for selected pain types. Panel B: Displays the association between the onset of  
240 widespread chronic pain, pain mechanism classification, and the proteomic scores (ProtS) in  
241 2019. Panels C and J: Show the associations between the onset of chronic pain and the  
242 proteomic scores in 2019 and 2022, respectively. "Onset of chronic pain" includes  
243 participants who were pain-free at baseline but reported chronic pain at follow-up. Panels D  
244 and K: "Pain spreading" refers to participants whose number of pain sites increased, while  
245 "pain recovery" refers to those whose pain sites decreased. Panels E and L: Depict the  
246 number of pain sites (multisite chronic pain) plotted against quantiles of the Sparse Proteomic  
247 Score (S-ProtS) and Intricate Proteomic Score (I-ProtS). Panel F: Shows the proportion of  
248 chronic pain duration across quantiles of I-ProtS. Panels J-I: Illustrate the rating of worst pain,  
249 impact of pain, and symptom severity, respectively, in relation to quantiles of S-ProtS and I-  
250 ProtS. Panel M: Illustrates the proportion of activity limitations caused by overall chronic  
251 pain among participants with chronic pain. Panel N: Shows activity limitations by specific  
252 chronic pain sites. Abbreviations: S-ProtS, Sparse Proteomic Score; I-ProtS, Intricate  
253 Proteomic Score.

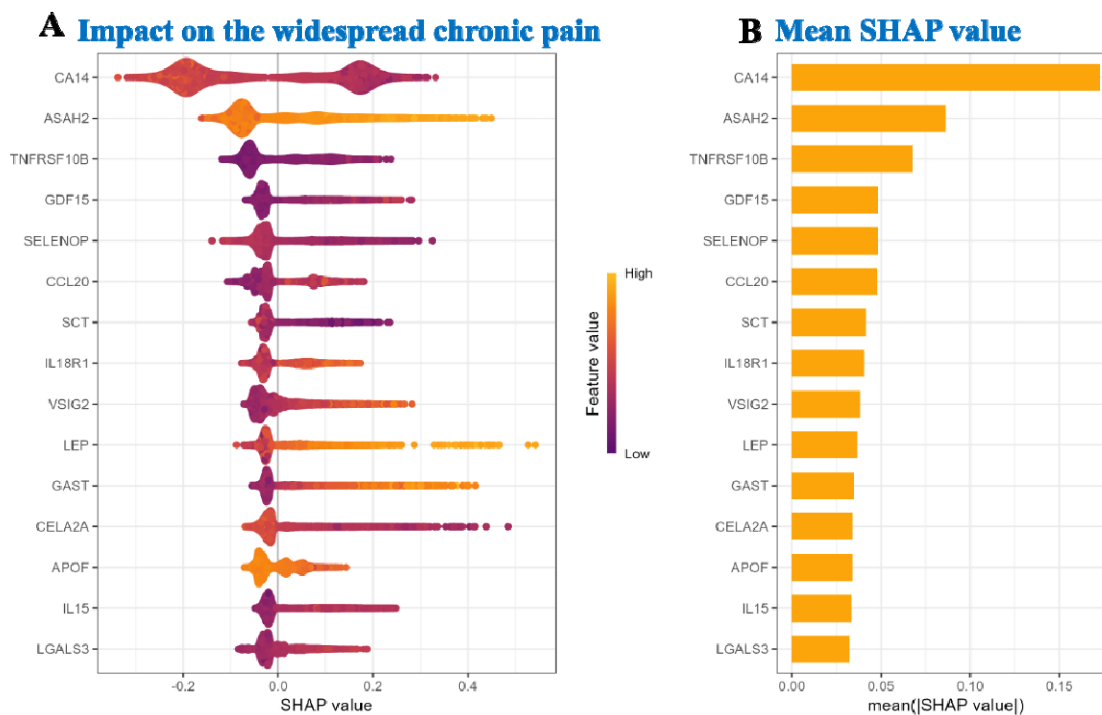
254

### 255 **Contributions of proteins to classification of widespread chronic pain**

256 Proteins including CA14, ASAH2, TNFRSF10B, GDF15, SELENOP, CCL20, SCT, IL18R1,  
257 VSIG2, LEP, GAST, CELA2A, APOF, IL15, and LGALS3 were identified as key  
258 contributors to widespread chronic pain. Among these, CA14 emerged as the most impactful,  
259 reflected by its highest SHAP (SHapley Additive exPlanations) value, suggesting a prominent  
260 role in pain prediction. ASAH2 and TNFRSF10B followed as major contributors. The SHAP  
261 analysis further highlighted the role of LEP, where both high and low levels influenced risk  
262 prediction. Elevated LEP values (indicated by yellow) are particularly associated with an  
263 increased risk of widespread chronic pain (**Fig. 5**).

264 SHAP analyses identified several key proteins contributing to the classification of different  
265 chronic pain types. LEP consistently emerged as the top predictor, exerting a strong influence  
266 on the model. Other proteins, including COL9A1, GAST, SELENOP, and CRT1, also  
267 displayed high SHAP values, underscoring their roles in pain prediction. COL9A1 and GAST  
268 demonstrated particularly notable impacts. In specific pain types, proteins such as  
269 TNFRSF10B, FETUB, SCT, and ITGAV also contributed significantly to prediction  
270 outcomes. GAST and FETUB ranked among the highest contributors in several subsets,  
271 while CA14, ASAH2, and PZP emerged as prominent predictors in others (**Extended Data**  
272 **Fig. 6**). Including participants with cancer did not change the results, as CA14 remained the  
273 top-ranking protein (**Extended Data Fig. 6**).

274



275

276 Fig. 5: SHAP of protein contributions to widespread chronic pain.

277 Notes: (A) SHAP plot illustrating the impact of individual proteins on widespread chronic  
278 pain prediction. Brighter colours (yellow) represent higher feature values, while darker  
279 colours (purple) indicate lower feature values, demonstrating how each protein contributes to  
280 the prediction model. (B) Mean SHAP values for the top contributing proteins to widespread  
281 chronic pain. Abbreviations: SHAP, SHapley Additive exPlanations.

282

### 283 **Prospective association of Top 15 proteins with pain phenotypes**

284 We investigated the prospective associations between the levels of the top 15 proteins of the  
285 widespread chronic pain, top 15 proteins (CA14, ASAH2, TNFRSF10B, GDF15, SELENOP,  
286 CCL20, SCT, IL18R1, VSIG2, LEP, GAST, CELA2A, APOF, IL15, and LGALS3), and  
287 mechanism pain classification (nociceptive pain, neuropathic pain, nociplastic pain, and  
288 fibromyalgia), the onset, spread, and recovery of chronic pain.

289 Except for CELA2A, the other proteins exhibited significant associations with nociplastic  
290 pain or fibromyalgia. None of the 15 proteins showed significant associations with  
291 nociceptive pain or neuropathic pain, indicating that these 14 proteins can effectively  
292 distinguish nociplastic pain/fibromyalgia from nociceptive and neuropathic pain up to 9 years  
293 in advance (**Fig.6A**). At the 2022 follow-up, the onset and progression of chronic pain  
294 showed significant associations with the top 11 proteins, excluding SCT, GAST, CELA2A,  
295 and IL15. Among these, CA14, TNFRSF10B, GDF15, VSIG2, and LEP exhibited significant  
296 associations with all aspects of chronic pain onset and progression, including pain spreading

297 and recovery (**Fig. 6A**). The detailed prospective associations between the top 15 proteins  
298 and pain mechanism classification (nociceptive pain, neuropathic pain, nociplastic pain, and  
299 fibromyalgia), as well as the onset and progression of chronic pain, including non-linear  
300 associations, are presented in the **Supplementary Fig.2-16**.

301

### 302 **Mendelian Randomisation and Causal Inferences**

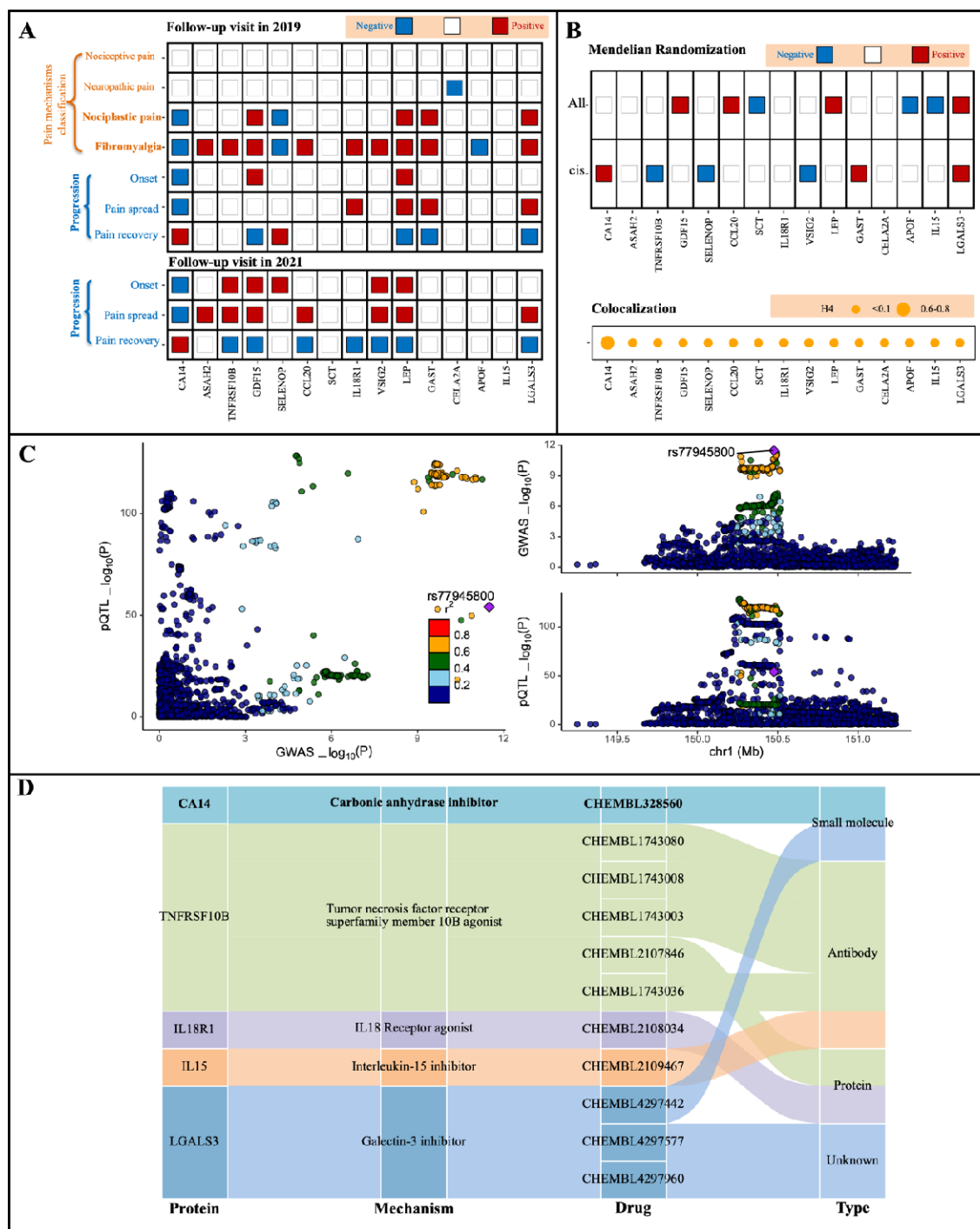
303 Two-sample Mendelian randomisation (MR) analyses revealed causal effects of genetically  
304 predicted levels of 12 proteins (all but *ASAH2*, *IL18R1*, and *CELA2A*) and multisite chronic  
305 pain (**Fig. 6B and Supplementary Table 5**). Further colocalisation analyses reinforced the  
306 causal relevance of *CA14* ( $H4 = 0.61$ ) with multisite chronic pain (**Fig. 6C and**  
307 **Supplementary Table 6**). In the reverse MR, we found no significant association between  
308 genetic liability to multisite chronic pain and *CA14*. However, a significant association was  
309 observed with *LEP* (**Supplementary Table 6**).

310 The protein-protein interaction (PPI) network revealed that *LEP*, *GDF15*, and *TNFRSF10B*  
311 were central nodes, exhibiting multiple connections with other significant proteins such as  
312 *LGALS3*, *IL15*, and *CCL20*, suggesting their pivotal roles in chronic pain modulation. In  
313 contrast, proteins like *CELA2A* and *APOF* were more isolated, indicating potentially  
314 peripheral roles in the network (**Extended Data Fig. 7 and Supplementary Table 7**).

315

### 316 **Proteomics-based drug repurposing**

317 Among the top 15 proteins, five (*CA14*, *TNFRSF10B*, *IL18R1*, *IL15*, and *LGALS3*) were  
318 found in the Open Targets database, highlighting their potential as therapeutic targets  
319 (Supplementary Table 8). A total of 11 drugs were identified targeting these proteins,  
320 including both small molecules and biologics such as antibodies and proteins, each  
321 categorised by drug type and mechanism of action. Notably, **CA14** emerged as a key  
322 therapeutic target, with sultiame (*CHEMBL328660*), a carbonic anhydrase inhibitor, as its  
323 main associated drug. *SULTHIAME*, which acts on the plasma membrane (**Extended Data**  
324 **Fig. 12**), is primarily used for treating obstructive sleep apnea and epilepsy. However, *CA14*  
325 agonists may be a more suitable therapeutic approach, based on the non-linear association of  
326 the *CA14*. Beyond sulthiame, the remaining ten drugs present promising options for treating  
327 nociplastic pain



328

329

Fig 6. Prospective associations, mendelian randomisation results, and drug repurposing

330

based on top 15 proteins and their potential causal relationships with chronic pain.

331

Notes: Panel A: prospective associations between the top 15 proteins and chronic pain. The

332

analysis was adjusted for several confounders, including sleeplessness, frustration, fatigue,

333

mood disturbances, stressful life events, and body mass index (BMI). Prognostic associations

334

for pain onset were estimated among participants who reported no pain at baseline. Panel B:

335

mendelian randomisation and colocalisation results for the top 15 proteins. the colocalisation

336

analysis utilised protein quantitative trait loci (pQTL) data from the UK biobank (UKB) to



337 identify shared genetic variants between protein levels and pain outcomes. Mendelian  
338 randomisation was conducted to explore potential causal relationships between the top 15  
339 proteins and multisite chronic pain. Panel C: colocalisation analysis for CA14 and multisite  
340 chronic pain. Panel D: identification of potential therapeutic targets based on the top 15  
341 proteins.  
342

## 343 **Discussion**

344 Our analysis revealed marked differences in the protein signatures of chronic versus acute  
345 pain. By leveraging these protein changes, we developed pain-specific proteomic scores that  
346 not only effectively discriminated individuals with chronic pain from those who were pain-  
347 free, but also predicted the intensity, phenotype, onset, spread, and recovery of chronic pain,  
348 particularly widespread chronic pain, a key feature of nociplastic pain. Notably, a sparse  
349 proteomic signature based on the top 10 proteins performed comparably to models based on  
350 conventional clinical variables alone. Among the top 15 proteins identified, 12 were  
351 significantly linked to these pain mechanisms, with 10 supported by Mendelian  
352 randomisation analysis. CA14, identified as the top-ranking protein and supported by both  
353 MR and colocalisation analyses, is a known target of sulthiame (CHEMBL328560),  
354 suggesting a potential therapeutic target for nociplastic pain conditions such as fibromyalgia.

### 355 **Distinct Protein Signatures in Chronic and Acute Pain**

356 Chronic pain exhibits more extensive protein changes compared to acute pain. Traditionally,  
357 chronic pain is defined primarily by its duration and lacks specific biomarkers to guide  
358 clinical management. Our study addresses this challenge by identifying distinct proteomic  
359 signatures that differentiate between chronic and acute pain, suggesting that proteomic  
360 information could predict those who will transition from acute to chronic pain. Previous  
361 research has highlighted the utility of proteomics in understanding pain sensitivity, especially  
362 in experimental models<sup>18</sup>. Notably, several proteins identified in our study provide novel  
363 insights into the biological pathways involved in pain mechanisms, in particular highlighting  
364 the involvement of metabolic and immune pathways in chronic pain development and  
365 persistence. For instance, leptin (LEP), known for its role in energy regulation, was strongly  
366 associated with chronic pain, indicating the potential involvement of metabolic dysregulation.  
367 Additionally, proteins such as ASGR1 and RARRES2, involved in immune responses,  
368 highlight the role of inflammation in chronic pain mechanisms. Inflammatory biomarkers  
369 may aid in predicting high-risk groups for developing nociplastic pain and guide the  
370 development of tailored treatments.

371

372 The distinct molecular profiles observed across specific pain types, such as head and face  
373 pain, abdominal pain, and musculoskeletal pain, further emphasise the heterogeneity of  
374 chronic pain. However, the identification of shared proteins across multiple pain types —  
375 such as APOF and ASGR1— suggests that there may be common underlying biological

376 mechanisms driving pain across body regions. This finding could pave the way for the  
377 development of universal biomarkers that apply across multiple pain syndromes.

378

### 379 **Classification Accuracy of Proteomic Profiling**

380 We demonstrate that proteomic profiling can perform as well as conventional clinical risk  
381 scores in predicting chronic pain and its progression. Moreover, combining proteomic scores  
382 with clinical variables further enhances predictive performance, suggesting that an integrated  
383 approach may be beneficial. This study is the first to comprehensively demonstrate the  
384 predictive power of plasma proteins across a broad spectrum of chronic pain types. We  
385 achieved this by directly comparing proteomic data with clinical variables and evaluating the  
386 added predictive value that proteomics offers beyond traditional clinical predictors. The  
387 proteomic score alone exhibited high predictive accuracy for chronic pain and its various  
388 subtypes. This predictive performance was further enhanced when proteomic data were  
389 combined with clinical variables. Our findings underscore the substantial potential of  
390 proteomic profiling as a promising technique to complement clinical evaluation, aiding in  
391 both the risk assessment for transition to chronic pain and the prognostication of chronic pain  
392 outcomes.

393

394 The proteomic score demonstrated a strong association with the onset, spread, and recovery  
395 of chronic pain. This profile effectively identifies individuals at higher risk for developing  
396 chronic pain, experiencing its spread to new sites, and their likelihood of recovery. Pain is  
397 influenced by a complex interplay of biological, psychological, and social factors<sup>13</sup>. Proteins  
398 integrate genetic, environmental, age-related, behavioural, and medication influences,  
399 offering comprehensive insights into pain mechanisms. Previous research has suggested that  
400 pain spread is not random but shows strong dependencies between adjacent pain sites<sup>13</sup>. Our  
401 study supports this by revealing significant associations between pain-specific proteomic  
402 scores and pain at various sites. Additionally, we observed robust correlations between  
403 proteomic scores and pain recovery, indicating shared biological foundations for both the  
404 spread and recovery of pain.

405

406 The proteomic score demonstrated a significant and specific association with nociplastic pain  
407 and fibromyalgia, distinguishing them from nociceptive and neuropathic pain. Nociplastic  
408 pain is characterised by abnormal sensory processing without clear peripheral damage. The  
409 association between proteomic score and nociplastic pain suggests that proteomic signatures

410 capture the complex biological underpinnings of this condition, including abnormalities in  
411 sensory perception, a hallmark feature<sup>11</sup>.

412

### 413 **Causal Relevance and Therapeutic Potential of Key Proteins**

414 Carbonic anhydrase 14 and leptin were identified as the primary proteins consistently  
415 associated with chronic pain, particularly nociplastic pain and fibromyalgia, and progression  
416 of chronic pain (onset, spreading, and recovery of chronic pain) in both longitudinal analyses.  
417 carbonic anhydrase 14 , identified here for the first time as associated with chronic pain,  
418 plays an important role in neuronal signal transmission via zinc ion binding and carbonate  
419 dehydratase activity. It is primarily enriched in the spinal cord and has a crucial role for neuronal  
420 function. Leptin is involved in immune regulation and inflammation, contributing to pain  
421 modulation, especially in nociplastic pain. Fibromyalgia affects 10–48% of patients with  
422 rheumatic diseases, compared to ~2–6% in the general population, likely due to shared  
423 mechanisms of immune dysregulation and inflammation<sup>19-21</sup>. Both leptin and gastrin showed  
424 positive correlations with new-onset pain and pain spread, suggesting their role in the  
425 development and propagation of chronic pain. This relationship is further supported by their  
426 impact on pain recovery outcomes. Animal studies have demonstrated that leptin can  
427 influence pain thresholds<sup>22,23</sup>. Together, these proteins suggest potential complementary roles  
428 in pain, with leptin influencing inflammation and carbonic anhydrase 14 affecting neuronal  
429 signalling.

430

431 The role of carbonic anhydrase 14 and leptin were both supported by mendelian  
432 randomisation, with carbonic anhydrase 14 further validated through colocalisation analysis.  
433 Our search on the Open Targets platform revealed that carbonic anhydrase 14 is the target of  
434 an approved small molecule drug, sulthiame, which is currently used to treat epilepsy and  
435 obstructive sleep apnoea. However, there are currently no well-documented drugs that  
436 activate CA14. Most research on carbonic anhydrase modulation has centered on inhibitors  
437 rather than activators, as seen with sulthiame in epilepsy treatment. If enhancing CA14  
438 activity proves to be therapeutically relevant for pain management, future research will need  
439 to focus on developing specific agonist or allosteric modulators. In addition to carbonic  
440 anhydrase 14, drugs for four other proteins were identified, primarily used for cancer and  
441 autoimmune diseases. While these therapies carry potential risks related to immune  
442 modulation, they remain possibilities for treatment options. Carbonic anhydrase 14, the top-

443 ranked protein with a strong causal association, suggests that targeting this protein could offer  
444 a promising therapeutic avenue for nociplastic pain.

#### 445 **Limitations and strengths**

446 While our study provides strong evidence for the role of proteomic profiling in  
447 characterising chronic pain, there are limitations which warrant discussion. First, while  
448 proteomic profiling demonstrated predictive value, the Olink panels used may have excluded  
449 key proteins relevant to chronic pain, potentially limiting the comprehensiveness of our  
450 analysis. Future studies should incorporate a wider array of proteomic markers to capture a  
451 more complete picture of the biological pathways involved in chronic pain. Second, the focus  
452 was primarily on a specific subset of pain types. Other pain characteristics not included may  
453 have different proteomic profiles and predictors. Furthermore, some participants in the  
454 control group, despite being classified as pain-free, may have had a history of chronic pain.  
455 Third, the study population was predominantly white British, which may limit the  
456 generalizability of the findings to other ethnic groups. Proteomic profiles and pain  
457 associations could vary across different genetic backgrounds. Fourth, while leptin and  
458 carbonic anhydrase 14 showed a causal relationship with pain, potential protein-protein  
459 interactions were not considered. The underlying biological pathways also remain unclear,  
460 warranting further research. Fifth, the findings have not been validated in independent  
461 cohorts. External validation in other large and diverse populations is necessary to confirm the  
462 robustness and applicability of the results. Finally, we acknowledge that the MR findings  
463 would be more robust if there were evidence indicating that the pQTLs and genetic variations  
464 associated with the outcome are shared. However, we also recognise that the lack of  
465 colocalisation evidence does not undermine the validity of the results, as colocalisation  
466 methods are known to have a high false-negative rate, typically around 60%<sup>24</sup>.

467

#### 468 **Conclusion**

469 In conclusion, this study underscores the significant potential of proteomic profiling to  
470 advance our understanding and management of chronic pain, particularly nociplastic pain and  
471 fibromyalgia. By identifying key proteins such as **carbonic anhydrase 14** and **leptin**, we  
472 offer new insights into biological mechanisms underlying pain. Therapies targeting  
473 TNFRSF10B, like tigatuzumab, represent promising candidates for drug repurposing in  
474 chronic pain treatment trials. Although no CA14 agonists are currently available, CA14  
475 remains a highly promising target, encouraging future efforts to develop novel therapies that  
476 leverage its potential in pain modulation.



## 478 **Methods**

### 479 **Study population**

480 The UK Biobank is a prospective, population-based cohort comprising over 500,000  
481 individuals aged 40-69, recruited from 22 assessment centers across the UK between 2006  
482 and 2010. In 2019, additional online surveys were conducted to track specific pain-related  
483 outcomes over time. A random sample of participants (n=53,029) were selected for the  
484 measurement of proteomics data and thus included for the subsequent analyses. More details  
485 of study design, cohort construction, and study window is provided in **Extended Data Fig. 8**  
486 and **Extended Data Fig. 9**. Ethical approval was obtained from the North West Multi-centre  
487 Research Ethics Committee (MREC, [https://www.ukbiobank.ac.uk/learn-more-about-uk-](https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics)  
488 [biobank/about-us/ethics](https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics)), with all participants providing written informed consent. This  
489 research was approved under application number 98,358.

490 To reduce confounding effects from cancer diagnosis or treatment, participants with cancer-  
491 related pain were excluded, allowing the focus to remain on chronic non-cancer pain.  
492 Additionally, participants selected by the UK Biobank Pharma Proteomics Project (UKB-PPP)  
493 Consortium were excluded from the analyses.

### 494 **Clinical variables**

495 Six variables capturing biological, psychological, and social factors were included as  
496 predictors for the construction of a clinical model. They were specified in prior with literature  
497 evidence suggesting their predictive value for chronic pain<sup>13</sup>. The factors were collected by  
498 an standardised touchscreen questionnaire in UK Biobank and include sleep difficulties  
499 (difficulty falling asleep, rated as quite a lot or very much), neuroticism (perceived life effort  
500 as constant, rated as quite a lot or very much), powerlessness (feelings of lack of energy or  
501 powerlessness, rated as quite a lot or very much), mental health issues (history of diagnosis or  
502 treatment by a doctor), life stressors (events such as divorce, death of a partner, or significant  
503 unemployment history), and a BMI over 30kg/m<sup>2</sup>. All the variables listed above were  
504 transformed into binary variables.

### 505 **Proteomic profiling**

506 The UK Biobank Pharma Proteomics Project consortium has generated extensive blood-  
507 based proteomic data. Proteomic profiling was conducted on EDTA-plasma samples from a  
508 cohort of 54,893 UKB participants at baseline recruitment, derived from a randomised subset  
509 representative of the broader UKB population. Utilising the Olink Explore 1536 and  
510 Expansion platforms, the profiling encompassed 2,923 unique proteins, assessed through  
511 2,941 assays. The protein targeting assays are organised into four 384-plex panels:

512 Inflammation, Oncology, Cardiometabolic, and Neurology. Detailed methodologies  
513 concerning assay execution, sample selection, and handling are available in the online  
514 document (<https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=1839>). No notable impacts  
515 from batch and plate variations, nor anomalies in protein coefficients of variation (CVs), were  
516 detected. The inter- and intraplate CVs for all OLINK panels remained below 20% and 10%,  
517 respectively, with a median protein CV of 6.7%. There were strong correlations observed for  
518 identical proteins across different panels and between the OLINK assay and independent  
519 assays within the UK Biobank. The proteomics data were provided as Normalised Protein  
520 eXpression (NPX) values on a  $\log_2$  transformation.

521 This analysis focuses on participants with less than 50% missing proteomic data. Additionally,  
522 three proteins missing in over 50% of participants were excluded. Post-quality control,  
523 missing NPX values were imputed using chained random forests, incorporating age and sex  
524 as predictors in the imputation process. The NPX data of proteins were standardised to have a  
525 mean of 0 and a variance of 1, following a standard normal distribution.

### 526 **Baseline pain phenotype**

527 At the baseline recruitment between 2006 and 2010, 501,243 participants were asked about  
528 pain experiences that interfered with their usual activities in eight major body regions in the  
529 past one month.

530 Participants who reported pain at a specific site in the past month were subsequently asked  
531 whether the pain had persisted for more than three months. This follow-up question was used  
532 to define chronic pain (pain persisting for more than three months) and acute pain (having  
533 pain in last one month but persisting for three months or less). Among the eight painful  
534 locations specifically asked in the questionnaire, we grouped them into four categories based  
535 on the similarity of underlying aetiologies and conditions: (1) headache and facial pain,  
536 which may reflect conditions like migraines or tension headaches; (2) abdominal pain,  
537 potentially linked to visceral causes such as irritable bowel syndrome (IBS) or endometriosis;  
538 (3) musculoskeletal pain (neck or shoulder pain, hip pain, and knee pain), which may indicate  
539 localized mechanical or inflammatory issues; and (4) widespread pain (“pain all over the  
540 body”), often associated with conditions like fibromyalgia. For widespread pain, participants  
541 who reported experiencing it were not given the option to select other specific body regions.

### 542 **Longitudinal Pain Phenotypes**

#### 543 **Online UK Biobank in 2019 - Pain mechanism classification:**

544 Pain outcomes were classified based on participants' chronic pain status and diagnostic  
545 criteria derived from validated scales. In 2019, participants were invited to complete an



546 online pain questionnaire that incorporated the **2016 revised Fibromyalgia Survey Criteria**  
547 **(FSC)**<sup>25</sup> and the **Douleur Neuropathique 4 (DN4)**<sup>26</sup>.

548 **Douleur Neuropathique 4 (DN4)**: The DN4 is a seven-item questionnaire designed to assess  
549 neuropathic pain. In the UK Biobank study, an abbreviated version was used, omitting the  
550 clinical examination assessment. Scores range from 0 to 7, with a score of  $\geq 3$  indicating the  
551 presence of neuropathic pain. Although primarily used to diagnose neuropathic pain, the DN4  
552 may also reflect central sensitisation, a key feature of nociplastic pain<sup>27</sup>.

553 **Fibromyalgia Survey Criteria (FSC)**: The FSC is based on the 2016 ACR Fibromyalgia  
554 Survey Criteria<sup>25</sup> and consists of two components: the **Widespread Pain Index (WPI)**,  
555 which measures pain across 19 body areas (score 0-19), and the **Symptom Severity Scale**  
556 **(SSS)**, which assesses somatic symptoms including fatigue, sleep disturbances, and cognitive  
557 difficulties (score 0-12). The total score (range 0-31) is calculated by combining the WPI and  
558 SSS scores.

559 The WPI was used to assess the severity of nociplastic pain, while the DN4 helped to identify  
560 both neuropathic pain and potential central sensitisation, providing a comprehensive  
561 evaluation of different pain types.

562

563 Based on these criteria, participants were classified into four distinct pain categories:

564 **Nociceptive pain** was defined for participants with chronic pain but no neuropathic  
565 characteristics (DN4 score = 0) and an FSC score  $< 3$ , indicating a low likelihood of  
566 nociplastic pain. Participants without chronic pain were classified as having no nociceptive  
567 pain.

568 **Neuropathic pain** was defined for participants with chronic pain and a DN4 score  $\geq 3$   
569 suggesting the presence of neuropathic pain. Participants with an FSC score  $\geq 3$  were  
570 excluded to reduce the likelihood of including those with possible nociplastic or  
571 fibromyalgia-related pain.

572 **Nociplastic pain** was defined for participants exhibiting DN4  $\geq 3$  or a total WPI and SSS  
573 score  $> 12$ . Notably, universally accepted diagnostic criteria for nociplastic pain remain absent.  
574 Fibromyalgia, widely recognized as the prototypical manifestation of nociplastic pain,  
575 highlights the utility of the FSC score as an effective marker of symptom severity. In addition,  
576 the FSC score may reliably reflect the burden of nociplastic pain<sup>28</sup>. Accordingly, an FSC  
577 score  $\geq 12$  serves as a practical threshold for identifying individuals likely to experience  
578 nociplastic pain as the predominant driver of their chronic pain.

579 **Fibromyalgia** was identified either through self-reported diagnoses from baseline or the  
580 2019 questionnaires or via general practitioner records. Additionally, participants with a WPI  
581 and SSS combined score >12 were classified as having fibromyalgia<sup>28</sup>.

582

583 **Online UK Biobank in 2019 - Pain sites:** Participants were inquired about experiencing  
584 pain or discomfort, either constantly or intermittently, for a duration exceeding three months.  
585 They were then requested to specify the affected sites, which included the head, face, neck,  
586 shoulders, chest, stomach, abdomen, back, hip, legs, knees, feet, arms, hands, and widespread  
587 body pain.

588 **New onset pain:** Participants who did not report any chronic pain at baseline but reported  
589 chronic pain at the follow-up visit were defined as having onset of pain.

590 **Spreading chronic pain:** The spreading of chronic pain was assessed by monitoring changes  
591 in the number of chronic pain sites. An increase in the number of pain sites at the follow-up  
592 visit compared to the baseline visit was defined as pain spread.

593 **Recovery of Chronic Pain:** Recovery of chronic pain was defined as a decrease in the  
594 number of pain sites at the follow-up visit compared to the baseline visit.

595 **Duration of Pain:** Participants reporting chronic pain were asked to specify the duration of  
596 their discomfort through the online questionnaire, selecting from the following categories: 3–  
597 12 months, 1–5 years, or more than 5 years.

598 **Pain Rating in the Last 24 Hours:** Participants were also asked to rate their pain intensity at  
599 each reported chronic pain site over the preceding 24 hours using a numeric rating scale of 0  
600 to 10, where 0 represented no pain and 10 indicated the worst possible pain.

601 **Worst Pain Rating:** Participants were prompted to rate the most severe pain they had  
602 experienced in the last 24 hours at the chronic pain site that bothered them the most, again on  
603 a numeric rating scale from 0 (no pain) to 10 (the worst imaginable pain). or widespread pain,  
604 a separate questionnaire ("All-over body pain in the last three months and rating of pain")  
605 was used. The ratings from specific body sites and the overall widespread pain were  
606 combined for analysis.

607 **Pain Interference:** The impact of pain on daily functioning was evaluated across seven  
608 domains: general activity, mood, walking ability, normal work, relationships with others,  
609 sleep, and enjoyment of life. Each domain was rated on a scale of 0 to 10, where 0 indicated  
610 no interference and 10 signified complete interference. A total pain interference score was  
611 derived by summing the individual scores for these seven items.

612 **Symptom Severity:** Participants were asked to report the severity of three symptoms—  
613 fatigue, sleep quality, and cognitive symptoms—over the past week, as measured by the  
614 fibromyalgia symptom severity scale. Each symptom was rated on a scale from 0 (no  
615 problem) to 3 (severe, pervasive, continuous, and life-disrupting issues). The total symptom  
616 severity score was calculated by summing the scores for fatigue, sleep quality, and cognitive  
617 symptoms.

618 **Online UK Biobank in 2022- Pain sites:** In 2022, the UK Biobank conducted an online  
619 survey to assess pain at various anatomical sites among participants. Respondents were asked  
620 to identify current pain in specific areas, including the neck, back, chest, abdomen, legs,  
621 muscles, and joints. Additionally, participants provided information on the duration of the  
622 pain and the degree to which each site of pain impacted their daily lives. Pain persisting for  
623 more than twelve weeks was classified as chronic pain.

624 The onset of chronic pain, spreading chronic pain, recovery of chronic pain in the follow-up  
625 questionnaire 2022 was also defined followed the phenotype at follow-up questionnaire 2019.

626

## 627 **Statistical analyses**

### 628 **Proteomic score**

629 We employed a rigorous four-step machine learning framework (**Extended Data Fig.10**),  
630 which included: (1) feature selection through protein-wide association studies (PWAS), (2)  
631 development of a XGBOOST (Extreme Gradient Boosting) model for intricate proteomic  
632 scores (I-ProtS) based on all proteins with significant associations after Bonferroni correction,  
633 (3) development of a XGBOOST for sparse proteomic scores (S-ProtS) based on the top 10  
634 proteins ranked by SHAP value, and (4) model validation. XGBoost is a high-performance  
635 machine learning algorithm designed for efficiency, flexibility, and portability, utilizing  
636 parallel tree boosting to develop the protein-profiling model<sup>29</sup>.

637

### 638 **Feature Selection**

639 All participants were included in the feature selection phase. A protein-wide association study  
640 was conducted on 2,923 proteins using a binomial regression model with Bonferroni  
641 correction. Proteins with p-values less than Bonferroni-corrected p-value (0.05/2923), were  
642 selected for the development of the GBM (gradient boosting machines) to generate the pain-  
643 specific proteomics scores.

644 The proteins identified by the protein-wide association study, associated with specific chronic  
645 pain, represent the protein signature. To explore these protein signatures in distinguishing

646 chronic from acute pain, acute pain — typically transient and driven by environmental factors  
647 — was used as a negative control to confirm the specificity of the protein signatures  
648 associated with chronic pain.

649

### 650 **Proteomic score development**

651 We developed a series of XGBOOST models to assess the potential of plasma proteomics as  
652 a single-domain assay for predicting widespread chronic pain. For each pain type, a separate  
653 XGBOOST model was constructed using 10-fold cross-validation, a technique employed to  
654 minimise the risk of random error caused by data splitting. The predicted risk values from  
655 these models were designated as proteomic scores (intricate proteomic scores, I-ProtS). To  
656 identify the most important proteins contributing to the prediction of each pain type, we used  
657 SHAP (SHapley Additive exPlanations) values. SHAP values are derived from cooperative  
658 game theory and provide a unified measure to explain the contribution of each feature—in  
659 this case, each protein—to the model’s prediction. A positive SHAP value indicates that the  
660 protein increases the predicted risk for widespread chronic pain, while a negative SHAP  
661 value suggests the protein decreases the predicted risk. For each pain type, the top 15 proteins  
662 were ranked based on their SHAP values, indicating their relative importance in the model.  
663 SHAP plots were also generated to visually interpret the impact of each protein on the  
664 model’s predictions, allowing for a clear understanding of how specific proteins influence the  
665 risk predictions for widespread chronic pain.

666 Based on the top 10 proteins, we developed an additional XGBOOST model to generate a  
667 spare-proteomic-score (S-ProtS), focusing on a more compact set of features. This model  
668 aims to reduce complexity while maintaining predictive power, leveraging only the most  
669 important proteins identified from the previous analysis.

### 670 **Model for pain prediction**

671 We fitted GBM models incorporating different sets of clinical predictors (**Extended Data**  
672 **Fig.11**). Specifically, 80% of participants were designated for training, while 20% were used  
673 for validation. We developed models based on six predictors identified from a previous study,  
674 utilizing 10-fold cross-validation. Additionally, we created models combining these six  
675 clinical predictors with proteomic score to evaluate their additive predictive value. Model  
676 performance was assessed using AUC, ranging from 0.5 (non-informative) to 1 (perfect  
677 discrimination). Calibration plots were employed to visualise the agreement between  
678 predicted risks and observed pain rates, while net benefit curves illustrated the additive  
679 predictive value of ProRS alongside different clinical predictor sets.

680 We reported key predictive metrics, including the area under the Receiver Operating  
681 Characteristic (ROC) curve (AUC). Calibration was evaluated using calibration plots and  
682 quantified through the calibration slope statistic, with a slope of 1 indicating optimal overall  
683 calibration. All models underwent development and evaluation through 10-fold cross-  
684 validation, with continuous variables standardised and categorical variables one-hot encoded.  
685 The results for all participants, including those with cancer, are provided as a sensitivity  
686 analysis in the supplementary materials to assess whether the presence of cancer affects the  
687 contribution of proteins.

#### 688 **Prospective association of the proteomic score**

689 To evaluate the prognostic association of the widespread chronic pain-specific proteomic  
690 score, Poisson regression was employed to estimate the relationship between the proteomic  
691 score (quantile 5 vs quantile 1) and nociceptive pain, neuropathic pain, nociplastic pain,  
692 fibromyalgia, onset chronic pain, pain spread, and pain recovery. For onset pain, the analysis  
693 focused on participants without any chronic pain at baseline who developed onset pain by the  
694 follow-up visit.

#### 695 **Prospective association of the top proteins**

696 The top 15 proteins were identified as those with significant contributions to widespread  
697 chronic pain. Poisson regression was employed to estimate the prognostic association  
698 between these top proteins and pain mechanism classification (nociceptive pain, neuropathic  
699 pain, nociplastic pain, and fibromyalgia), new onset chronic pain, pain spread, and pain  
700 recovery (quantile 5 vs quantile 1). The analysis was adjusted for variables including  
701 sleeplessness, feeling fed-up, tiredness, mood, stressful life events, and BMI.

702 To explore dose-response relationships between the top proteins and pain outcomes, restricted  
703 cubic spline (RCS) regression was used, with the same multivariable adjustments.  
704 Nonlinearity was tested using the likelihood ratio test, and the number of nodes was selected  
705 based on the lowest Akaike Information Criterion (AIC) value. The same analytical approach  
706 was applied to the other top 15 proteins.

707

#### 708 **Causal relevance of the top proteins (Mendelian Randomisation Analysis)**

709 To explore the causal relationship between specific proteins and pain, we conducted a two-  
710 sample Mendelian randomisation (MR) analysis. This study targeted the top 2 shared proteins  
711 frequently appearing among the top ten across pain types: leptin (LEP) and gastrin (GAST).

712 Selection of genetic instruments: Instrumental variables were derived from two large-scale  
713 genome-wide association studies (GWAS): the UK Biobank Pharma Proteomics Project

714 (UKB-PPP). 1. All SNPs MR Analysis: Single nucleotide polymorphisms (SNPs)  
715 significantly associated with the expression of specific proteins in blood were selected as  
716 instrumental variables ( $p < 5 \times 10^{-5}$ ). To minimise the influence of strong linkage  
717 disequilibrium (LD) on the outcomes, SNPs with low weak LD ( $R^2 < 0.1$ ) within a 10000 kb  
718 window, as estimated using the 1000 Genomes European panel, were included. Additionally,  
719 F-statistics  $> 10$  were applied to ensure sufficient instrumental strength related to exposure  
720 traits. 2. Cis-SNPs MR Analysis: Cis-SNPs associated with the top shared plasma proteins at  
721 genome-wide significance ( $p < 5 \times 10^{-5}$ ) and LD ( $R^2 < 0.1$ ) were utilised as instrumental  
722 variables. Cis-SNPs were defined as those located within 5 Mb of the gene encoding the  
723 protein, with LD estimated based on the 1000 Genomes European panel. Furthermore, F-  
724 statistics  $> 10$  were applied to ensure adequate instrumental strength associated with exposure  
725 traits. This analysis was considered the primary MR result due to its proximal relevance to  
726 the gene encoding the protein.

727 Reverse MR analysis: The reverse MR used SNPs associated with multisite chronic pain at  $p$   
728  $< 5 \times 10^{-8}$ . Only SNPs with  $R^2 < 0.001$  within a 10,000 kb window were included, and all  
729 instruments had an F-statistic  $> 10$ .

730 Protein source: UKB-PPP.

731 Pain sources: To avoid potential bias arising from population diversity, we restricted the  
732 genetic background in our drug target MR analysis to participants of European descent. We  
733 extracted multisite chronic pain data from a large cohort of GWAS<sup>30</sup>, which included 73,082  
734 cases and 105,474 controls from the latest UK Biobank datasets.

### 735 **Sensitivity analysis**

736 The Steiger test was employed to assess potential reverse causality in the MR analysis.  
737 Additionally, the bidirectional Mendelian randomisation approach was used to explore the  
738 reverse causal relationship between multisite chronic pain and specific proteins. Associations  
739 between cis-pQTLs and outcomes were calculated as beta with corresponding confidence  
740 intervals (CIs) using the random-effects inverse-variance weighted (IVW) method.  
741 Furthermore, MR-PRESSO was utilised to detect outlier genetic variants. The MR-Egger  
742 intercept test was conducted to evaluate horizontal pleiotropy, and Cochran's Q value was  
743 calculated to assess heterogeneity in IVW estimators, with  $P < 0.05$  indicating evidence of  
744 heterogeneity.

### 745 **Colocalisation analysis**

746 We performed colocalisation analysis to determine whether the observed associations  
747 between the top shared proteins and multisite chronic pain were driven by linkage

748 disequilibrium. For each locus, a Bayesian framework was applied to evaluate support for  
749 five mutually exclusive hypotheses: (1) no association with either trait; (2) association with  
750 trait 1 only; (3) association with trait 2 only; (4) both traits are associated but have distinct  
751 causal variants; and (5) both traits are associated and share the same causal variant. The  
752 analysis yielded posterior probabilities for each hypothesis (H0, H1, H2, H3, and H4). We  
753 assigned prior probabilities as follows: the SNP being associated with trait 1 only ( $p_1$ ) was set  
754 at  $1 \times 10^{-4}$ ; the SNP being associated with trait 2 only ( $p_2$ ) at  $1 \times 10^{-4}$ ; and the SNP being  
755 associated with both traits ( $p_{12}$ ) at  $1 \times 10^{-4}$ . Strong evidence of colocalisation was defined  
756 as a posterior probability for shared causal variants ( $PH_4$ )  $\geq 0.8$ , while medium evidence was  
757 indicated by  $0.5 < PH_4 < 0.8$ .

### 758 **Protein-protein interaction (PPI)**

759 We performed protein-protein interaction (PPI) analysis for the top 15 proteins identified in  
760 our models using the **STRING database** (<https://string-db.org/>)<sup>31</sup>. STRING offers  
761 comprehensive interaction networks derived from experimental data, computational  
762 predictions, and known biological pathways. This analysis enabled us to examine potential  
763 interactions between the top proteins and to identify biologically relevant networks that may  
764 be involved in the pathogenesis of widespread chronic pain.

### 765 **Drug Repurposing**

766 To explore the potential for drug repurposing, we searched the Open Targets platform<sup>32</sup> for  
767 the top 15 proteins identified in our analysis. Open Targets integrates data from multiple  
768 sources, including genetic associations, known drug interactions, and disease pathways, to  
769 evaluate the therapeutic potential of various proteins. This search enabled us to determine  
770 whether any of the top proteins were associated with existing drugs that could be repurposed  
771 for chronic pain treatment. By identifying proteins already targeted by approved or  
772 investigational drugs, we aimed to uncover opportunities for repurposing these drugs to  
773 address widespread chronic pain. This approach may offer faster and more cost-effective  
774 therapeutic options by leveraging drugs that have already undergone substantial testing for  
775 safety and efficacy in other conditions.

776

777 All modeling and statistical analyses were performed using R software (V4.3.0). Missing  
778 value imputation was conducted using the missRanger package. The XGBOOST and gradient  
779 boosting model (GBM) was implemented with the caret package, with calibration curves  
780 generated by the CalibrationCurve package and net benefit curves by the rmda package. The

781 two-sample Mendelian randomisation analysis was conducted using the TwosampleMR  
782 package. Colocalisation analysis was carried out with the *coloc* package.



783

784 Reference

- 785 1. Zimmer, Z., Fraser, K., Grol-Prokopczyk, H. & Zajacova, A. A global study of pain  
786 prevalence across 52 countries: examining the role of country-level contextual factors.  
787 *Pain* **163**, 1740-1750 (2022).
- 788 2. Tompkins, D.A., Hobelmann, J.G. & Compton, P. Providing chronic pain  
789 management in the "Fifth Vital Sign" Era: Historical and treatment perspectives on a  
790 modern-day medical dilemma. *Drug Alcohol Depend* **173 Suppl 1**, S11-S21 (2017).
- 791 3. Pleis, J.R., Ward, B.W. & Lucas, J.W. Summary health statistics for U.S. adults:  
792 National Health Interview Survey, 2009. *Vital Health Stat* **10**, 1-207 (2010).
- 793 4. Husky, M.M., Ferdous Farin, F., Compagnone, P., Fermanian, C. & Kovess-Masfety,  
794 V. Chronic back pain and its association with quality of life in a large French  
795 population survey. *Health Qual Life Outcomes* **16**, 195 (2018).
- 796 5. Soni, A., *et al.* Hospitalization in fibromyalgia: a cohort-level observational study of  
797 in-patient procedures, costs and geographical variation in England. *Rheumatology*  
798 (*Oxford*) **59**, 2074-2084 (2020).
- 799 6. Phillips, C.J. The Cost and Burden of Chronic Pain. *Rev Pain* **3**, 2-5 (2009).
- 800 7. Raja, S.N., *et al.* The revised International Association for the Study of Pain definition  
801 of pain: concepts, challenges, and compromises. *Pain* **161**, 1976-1982 (2020).
- 802 8. Treede, R.D., *et al.* A classification of chronic pain for ICD-11. *Pain* **156**, 1003-1007  
803 (2015).
- 804 9. Treede, R.D., *et al.* Chronic pain as a symptom or a disease: the IASP Classification  
805 of Chronic Pain for the International Classification of Diseases (ICD-11). *Pain* **160**,  
806 19-27 (2019).
- 807 10. Gatchel, R.J., Peng, Y.B., Peters, M.L., Fuchs, P.N. & Turk, D.C. The biopsychosocial  
808 approach to chronic pain: scientific advances and future directions. *Psychol Bull* **133**,  
809 581-624 (2007).
- 810 11. Kaplan, C.M., *et al.* Deciphering nociplastic pain: clinical features, risk factors and  
811 potential mechanisms. *Nat Rev Neurol* **20**, 347-363 (2024).
- 812 12. Carrasco-Zanini, J., *et al.* Proteomic signatures improve risk prediction for common  
813 and rare diseases. *Nat Med* **30**, 2489-2498 (2024).
- 814 13. Tanguay-Sabourin, C., *et al.* A prognostic risk score for development and spread of  
815 chronic pain. *Nat Med* **29**, 1821-1831 (2023).
- 816 14. Nielsen, C.S., Knudsen, G.P. & Steingrimsdottir, O.A. Twin studies of pain. *Clin*  
817 *Genet* **82**, 331-340 (2012).
- 818 15. Abboud, C., *et al.* Animal models of pain: Diversity and benefits. *J Neurosci Methods*  
819 **348**, 108997 (2021).
- 820 16. Gottesman-Katz, L., Latorre, R., Vanner, S., Schmidt, B.L. & Bunnett, N.W. Targeting  
821 G protein-coupled receptors for the treatment of chronic pain in the digestive system.  
822 *Gut* **70**, 970-981 (2021).
- 823 17. Denk, F., McMahon, S.B. & Tracey, I. Pain vulnerability: a neurobiological  
824 perspective. *Nat Neurosci* **17**, 192-200 (2014).
- 825 18. Segelcke, D., *et al.* BLOOD PROTEOMICS AND PAIN - A TRANSLATIONAL  
826 STUDY TO PROGNOSTICATE PAIN PHENOTYPES AND ASSESS NEW  
827 BIOMARKERS FOR PREVENTING PAIN IN HUMANS.  
828 2024.2007.2004.24309933 (2024).
- 829 19. Wolfe, F., Ross, K., Anderson, J., Russell, I.J. & Hebert, L. The prevalence and  
830 characteristics of fibromyalgia in the general population. *Arthritis Rheum* **38**, 19-28  
831 (1995).

- 832 20. Walitt, B., Nahin, R.L., Katz, R.S., Bergman, M.J. & Wolfe, F. The Prevalence and  
833 Characteristics of Fibromyalgia in the 2012 National Health Interview Survey. *PLoS*  
834 *One* **10**, e0138024 (2015).
- 835 21. Murphy, A.E., Minhas, D., Clauw, D.J. & Lee, Y.C. Identifying and Managing  
836 Nociceptive Pain in Individuals With Rheumatic Diseases: A Narrative Review.  
837 *Arthritis Care Res (Hoboken)* **75**, 2215-2222 (2023).
- 838 22. Tian, Y., *et al.* Leptin enhances NMDA-induced spinal excitation in rats: A functional  
839 link between adipocytokine and neuropathic pain. *Pain* **152**, 1263-1271 (2011).
- 840 23. Rodgers, H.M., Liban, S. & Wilson, L.M. Attenuated pain response of obese mice  
841 (B6.Cg-lep(ob)) is affected by aging and leptin but not sex. *Physiol Behav* **123**, 80-85  
842 (2014).
- 843 24. Hukku, A., *et al.* Probabilistic colocalization of genetic variants from complex and  
844 molecular traits: promise and limitations. *Am J Hum Genet* **108**, 25-35 (2021).
- 845 25. Wolfe, F., Egloff, N. & Hauser, W. Widespread Pain and Low Widespread Pain Index  
846 Scores among Fibromyalgia-positive Cases Assessed with the 2010/2011  
847 Fibromyalgia Criteria. *J Rheumatol* **43**, 1743-1748 (2016).
- 848 26. Bouhassira, D., *et al.* Comparison of pain syndromes associated with nervous or  
849 somatic lesions and development of a new neuropathic pain diagnostic questionnaire  
850 (DN4). *Pain* **114**, 29-36 (2005).
- 851 27. Rutter-Locher, Z., *et al.* A systematic review and meta-analysis of questionnaires to  
852 screen for pain sensitisation and neuropathic like pain in inflammatory arthritis. *Semin*  
853 *Arthritis Rheum* **61**, 152207 (2023).
- 854 28. Wolfe, F., Walitt, B.T., Rasker, J.J., Katz, R.S. & Hauser, W. The Use of  
855 Polysymptomatic Distress Categories in the Evaluation of Fibromyalgia (FM) and FM  
856 Severity. *J Rheumatol* **42**, 1494-1501 (2015).
- 857 29. Xue, T., *et al.* Extracellular vesicle biomarkers for complement dysfunction in  
858 schizophrenia. *Brain* **147**, 1075-1086 (2024).
- 859 30. Johnston, K.J.A., *et al.* Genome-wide association study of multisite chronic pain in  
860 UK Biobank. *PLoS Genet* **15**, e1008164 (2019).
- 861 31. Szklarczyk, D., *et al.* The STRING database in 2023: protein-protein association  
862 networks and functional enrichment analyses for any sequenced genome of interest.  
863 *Nucleic Acids Res* **51**, D638-D646 (2023).
- 864 32. Ochoa, D., *et al.* Open Targets Platform: supporting systematic drug-target  
865 identification and prioritisation. *Nucleic Acids Res* **49**, D1302-D1310 (2021).
- 866
- 867