ARTICLE IN PRESS

Nature Communications

https://doi.org/10.1038/s41467-026-71434-8

Article in Press

GWAS of extended prescription analgesic use identifies genetic loci in chronic pain

Received: 6 February 2025

Accepted: 3 March 2026

Cite this article as: Harlow, C.E., Uzochukwu, E., Fernando, H.A. *et al.* GWAS of extended prescription analgesic use identifies genetic loci in chronic pain. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-71434-8

Charli E. Harlow, Emeka Uzochukwu, Hazel A. Fernando, Charles E. Mordaunt, Jordan M. Hughey, John D. Eicher, Lara Robinson, Nicholas Bowker, Laurence Howe, Jimmy Liu, Adrian Cortes, Paul Wilson, Usha Gungabissoon, Victoria S. Benson, Anthony Nash, Gareth Young, Laura Addis, Chun-Fang Xu, Caleb Webber, Jonathan Davitte & M. Zameel Cader

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nd/4.0/.

ARTICLE IN PRESS

Title

GWAS of Extended Prescription Analgesic Use Identifies Genetic Loci in Chronic Pain.

Author List

Charli E. Harlow[1*], Emeka Uzochukwu[2], Hazel A. Fernando[1,3], Charles E. Mordaunt[4], Jordan M. Hughey[4], John D. Eicher[5], Lara Robinson[6], Nicholas Bowker[1], Laurence Howe[1], Jimmy Liu[4], Adrian Cortes[7], Paul Wilson[1], Usha Gungabissoon[8], Vicky Tribble[1], Anthony Nash[6], Gareth Young[4], Laura Addis[1], Chun-Fang Xu[1], Caleb Webber[2], Jonathan Davitte[4+], M. Zameel Cader[6+]

Affiliations

1. GlaxoSmithKline, Gunnels Wood Road, Stevenage, UK
2. UK Dementia Research Institute, Department of Psychological Medicine and Clinical Neuroscience, University of Cardiff, UK
3. Leeds Institute of Data Analytics, University of Leeds, UK
4. GlaxoSmithKline, Upper Providence, USA
5. GlaxoSmithKline, Cambridge, USA
6. Nuffield Department of Clinical Neuroscience, Kavli Institute for Nanoscience Discovery, Dorothy Crowfoot Hodgkin Building, Sherrington Road, University of Oxford, UK
7. GlaxoSmithKline, Heidelberg, Germany
8. GlaxoSmithKline, Oxford Street, London, UK

+ indicates joint senior authors

***Corresponding authors: Charli Harlow (charli.e.harlow@gsk.com); Jonathan Davitte (jonathan.m.davitte@gsk.com)**

Abstract (145 Words)

Pain-related conditions are the leading cause of disability worldwide. Existing GWAS for chronic pain have mainly focused on individual pain-related disorders, which may not optimally capture the phenotype. Here, we define chronic pain based on prescription analgesic use (≥90 days) in two large biobanks (UK Biobank and FinnGen). GWAS meta-analyses of 11 prescription-based pain phenotypes identify 140 associations with chronic pain, including 78 novel (e.g. ARPP21, CNTNAP2) and 62 previously reported (e.g. SLC39A8, DCC, TRPM8) associations. Integrating these genetic associations with functional data including transcriptome-wide association studies, cell-type and pathway enrichment, and gene enrichment in mouse phenotypes identifies potential mechanisms involved in chronic pain, implicating oligodendrocyte differentiation, neuronal guidance, endolysosomal function and post-synaptic endosome recycling. Our study showcases how the use of prescription data to identify and characterize pain can provide insights into pain genetics and its underlying biology.

Introduction

Pain-related conditions are the leading cause of disability worldwide (1). Chronic pain is defined as pain that persists or recurs for 3 months or longer (2). It encompasses a wide range of disorders with complex aetiologies from chronic headache to fibromyalgia. Chronic pain has been reported to affect 13–50% of adults in the UK; and among this group, 10.4–14.3% have moderate-to-severe disabling chronic pain (3).

ARTICLE IN PRESS

The current range of pharmacological treatments for chronic pain, including opiates, can be associated with harmful effects (4, 5) and/or have limited long-term effectiveness (6). Given the significant unmet medical need for individuals in chronic pain and challenges with existing treatments, there is an urgent need to identify targets linked to chronic pain using human data to allow the development of therapies for improved long-term outcomes.

To date, most genetic studies of pain have focused on the study of individual pain-related disorders (e.g. migraine, osteoarthritis, back pain). Identified common variants have implicated genes in neurological pathways and inflammation (7). However, focusing analyses on the genetic susceptibility to individual pain-related disorders limits the sample size for studying 'chronic pain' in general; and, consequently may miss common biology shared across multiple pain-related diseases. Recent genetic studies leveraging 'pain scores' based on patient reports or combining pain traits to increase sample size have led to the identification of genetic loci associated with pain (8-12). Pain scores are subjective and focus on the genetics underlying pain severity and individual pain perception. The ambiguity and heterogeneity in how patients score their pain makes interpretation of these results and their role in the aetiology of pain difficult. The investigation of non-single-chronic pain conditions by combining pain phenotypes identified using self-report or diagnostic codes is objective providing insights into the underlying disease aetiology cause of pain. However, combining multiple pain traits may result in spurious pleiotropy and make for difficult interpretation. Additionally, the use of self-report may lack phenotype depth or knowledge of pain-specific information, and may introduce biases.

While leveraging diagnosis codes to identify specific pain-related case populations for genetic characterisation is quite straight-forward in its implementation, many individuals never receive a specific pain-related diagnosis and are missed by the diagnosis-focused approach. These individuals may nevertheless seek medical care to repeatedly obtain prescription analgesics to control their chronic pain. Consequently, the availability of prescription data linked to genetic samples in large biobank-data, such as the UK Biobank and FinnGen, provides a unique opportunity for defining chronic pain. In this case, ascertainment is not based on the diagnosis of specific pain-related aetiologies, but rather on the identification of individuals that are seeking treatment for pain relief for extended periods of time. Extended use of analgesics may thus serve as a way of defining chronic pain that: 1) captures both the diagnosed and undiagnosed chronic pain population; 2) increases the total sample size for genetic characterisation of chronic pain; and, 3) aids in the identification of pain loci and an improved understanding of underlying biological pathways or mechanisms that are shared across pain-related aetiologies. Song *et al.* (13) recently performed a GWAS on the number of times codeine was prescribed highlighting how analgesic data can be interrogated to identify genes associated with pain. Here, we aim to build upon this approach utilising a broader list of pain medications to define a comprehensive chronic pain phenotype.

We used a unique operational definition of chronic pain (analgesic use $\geq$90 days) and conducted genome-wide association studies (GWAS) of pain phenotypes derived from prescription data in individuals of European ancestry in both the UK Biobank (N~500,000) and FinnGen (N~500,000). We subsequently completed meta-analyses of the results from these two biobanks to identify common variants associated with these pain phenotypes. We also conducted rare-variant gene-based tests, enabled by the Whole-Genome Sequencing (WGS) data in UK Biobank, to identify rare variants associated with these traits. Additionally, we sought to understand the genetic correlation between our definitions of chronic pain with other pain-related aetiologies and psychological disorders (e.g. bipolar or depression). Finally, we integrated all our results with genomics analyses from disease-relevant tissues, including gene expression, protein abundance, mouse knockout models, and human disease data, to

identify likely effector genes and enhance our understanding of the underlying biological pathways and disease aetiology.

Results

Discovery of 140 genetic associations with chronic pain

To increase our understanding of chronic pain, we used prescription data to derive 11 chronic pain phenotypes encompassing pain susceptibility, pain severity based on strong opioid use, and pain duration based on time on analgesics using UK Biobank and FinnGen (Figure 1, Supplementary Data 1). We performed GWAS on each of these phenotypes within each dataset followed by meta-analysis of each phenotype (overall sample sizes: 23,547 – 692,039; Figure 1, Supplementary Data 2). These combined meta-analyses identified 140 genome-wide associations ($P < 5 \times 10^{-8}$), of which 78 have not previously been reported as associated with pain phenotypes (Figure 2, Table 1, Supplementary Data 3). We detected many associations at loci previously reported in the literature (e.g., *SLC39A8, TRPM8)*, demonstrating that our prescription-based definitions of chronic pain are capturing relevant disease information (Table 1, Supplementary Data 3, Supplementary Data 4) (7, 10, 11, 14-16). Notably, 11 loci previously linked to opioid use disorder and drug metabolism showed association with our prescription defined pain traits (e.g., *SGIP1/PDE4B*, *OPRM1*, *CYP2D6/7*) (17, 18) (Supplementary Data 4).

We determined whether the use of prescription data highlighted associations linked to specific pain phenotypes (i.e., susceptibility to chronic pain vs severity vs duration) or were shared across multiple phenotypic domains by assessing the overlap of associations across the three phenotypic definitions (susceptibility, severity and duration). Most loci showed association with only chronic pain susceptibility phenotypes, indicating our phenotype definition is predominantly identifying genes related to central pain mechanisms (Figure 2A, Supplementary Data 3). When using strong opioid use as a measurement of pain severity, we identified two loci—*PAX1* and *KLHL29* (Figure 2A, Supplementary Data 3). The *PAX1* locus was only associated at genome-wide significant levels ($P < 5 \times 10^{-8}$) with the pain severity phenotypes, indicating that it may play a role in pain perception, whilst the *KLHL29* locus was also associated with pain susceptibility highlighting a potential function in both influencing central pain mechanisms and sensitisation. The five genetic loci (*NCKAP5, IGF2BP3, ETS1/KIRREL3, SPPL3, RIMBP2*) associated with pain duration were not associated with pain severity or pain susceptibility at genome-wide significant levels implicating their potential roles in the more temporal aspects of pain (Figure 2A, Supplementary Data 3).

ARTICLE IN PRESS

Sex-stratified analysis identified female- and male-specific GWAS loci for chronic pain

There are noted sex differences in pain perception and treatment with women often experiencing increased pain sensitivity, pain intensity, and recommendation to use opioid treatments (19). We performed sex-stratified GWASs for the 11 derived phenotypes in each cohort, meta-analysed each phenotype, and identified 87 GWAS associations, of which 11 and 21 genome-wide associations were only seen in males or females, respectively (Figure 2B, Supplementary Data 5, Supplementary Data 6, Supplementary Data 7). The slightly larger sample sizes in females compared to males may partially explain why we see a larger number of female-specific loci (Supplementary Data 2). Similar to the analyses not stratified by sex, the majority of the loci identified in the sex-stratified analyses (n=32 loci) were associated with chronic pain susceptibility. However, for the pain severity phenotypes, we noted genome-wide significant associations ($P < 5 \times 10^{-8}$) at three loci in males only, potentially suggesting that these genes may play a factor in influencing sex-specific differences in use of strong opioids in pain patients (Figure 2C, Supplementary Data 6, Supplementary Data 7).

ARTICLE IN PRESS

Prioritisation of candidate genes for the GWAS associations

To prioritise genes most likely driving the GWAS associations detected in our analyses, we employed a series of complementary computational approaches. First, we annotated the nearest gene based on the distance of the lead genetic variant to the transcription start site or gene-body and found that the majority of GWAS loci mapped to protein-coding genes (Supplementary Data 4). Next, we determined whether a coding variant with functional impact (i.e., moderate or high impact annotation using the Variant Effect Prediction (VEP) tool) was in the credible set of the GWAS signal with a posterior probability > 0.01. Functional variants impacting 7 known pain genes (*TRPM8, SEMA3F, ANAPC4, SLC39A8, MKRN1, NUCB2, C17orf58, CYP2D6,* and *CYP2D7*) and 4 genes not previously linked to pain (*ALK*, *NPC1, CD6,* and *NAA38)* were identified, implicating these genes as the likely candidate genes for these GWAS signals (Supplementary Data 4). Finally, expression and protein quantitative trait loci (eQTL/pQTL) colocalizations (posterior probability > 0.8 & QTL signal P-value < 1×10^{-5}) were performed to prioritise genes within GWAS loci that may influence phenotypes through modulation of gene and/or protein expression. At 20 loci, there was evidence of eQTL and/or pQTL colocalization for a single gene implicating this gene as the most likely causal gene within the locus (e.g., *NCAM1, ALK, LRP1)* (Supplementary Data 4). Prioritisation of the most likely candidate gene was more complex at 37 GWAS loci where there was e/pQTL colocalization for multiple genes within the locus (Supplementary Data 4). We applied a similar prioritisation workflow in the sex-stratified analyses highlighting functional variants impacting 8 genes (*SEMA3F, UHRF1BP1, NUCB2, C17orf58, CYP2D6, CYP2D7*, *PER3*, and *NLRC5*) and evidence of colocalization for only one gene at 8 independent loci (e.g. *OPRM1, ARPP21)* (Supplementary Data 7).

Functional analysis highlights key cell-types and gene-sets linked to chronic pain

We performed a series of functional analyses integrating orthogonal data types (e.g., cell-type specific expression, pathway enrichment, and protein-protein interaction (PPI) networks) to better understand the underlying biology and pathophysiology of pain. We integrated our genetic findings with cell-type gene expression data from cortex, substantia nigra (SN) (20), trigeminal ganglia (TG) (21), dorsal root ganglia (DRG) (22), and gut (22, 23), identifying 12 cell-types associated with chronic pain or migraine (Figure 3A, Supplementary Data 8). The strongest association was seen between cell-type-specific genes from Oligodendrocyte Progenitor Cells (OPCs) in the cortex with pain susceptibility (strong opioid use vs healthy controls: $P = 7 \times 10^{-4}$) (Figure 3A, Supplementary Data 8, Supplementary Figure 3). Other pain susceptibility associations were with neurons of the gut and peptidergic neurons (Figure 3A, Supplementary Data 8, Supplementary Figure 3).

We next formed PPI modules consisting of the OPC specific expressed genes and re-tested the association between pain phenotypes and OPC specific PPI modules to functionally define the association between pain susceptibility and OPC specific expressed genes. We found that two modules were significantly associated—one functionally annotated as associated with protein localisation (Module 4 $P = 9 \times 10^{-22}$, Figure 3B, Supplementary Data 9) and the other annotated as associated with cell-fate commitment and regulation of cell differentiation (Module 5 $P = 1.1 \times 10^{-10}$ and $P = 5.3 \times 10^{-15}$, respectively, Figure 3B, Supplementary Data 9). Network analysis of the genes present in cell fate commitment and regulation of cell differentiation highlighted genes previously associated with pain disorders (e.g. *HDAC1, NOTCH2, HES1, SOX5*) as well as genes not previously associated with pain disorders (e.g. *SOX2, SOX13, ASCL1*) (Figure 3C, Supplementary Data 10).

ARTICLE IN PRESS

Functional enrichment analysis with mouse knockout model phenotypes suggests genes, cell-types, and pathways related to chronic pain

To further delineate the underlying pathophysiology of pain and how pain genes may function *in vivo*, we utilised mouse phenotype gene-sets where genetic perturbations are mapped to *in vivo* phenotypes, with a total of 192,000 gene-phenotype pairs (24). We tested for enrichment of these mouse phenotype gene-sets in cell-type-specific genes from 43 cell-types across five tissues (described above—cortex, SN, TG, DRG, and gut). We found significant enrichment for 666 unique mouse phenotype gene-sets (Supplementary Data 11); 403 unique phenotypes are observed in the 12 cell-types we had linked to pain GWASs (Figure 3D). Significantly enriched phenotypes tended to align to the cell-type's function (e.g., neurological phenotypes associated with neurons and immune phenotypes associated with B cells).

We next tested whether our chronic pain GWASs were enriched for these identified cell-type specific mouse phenotype genes. We found the "strong opioid use" GWAS to be significantly enriched with peptidergic nociceptor-specific genes mapped to bilateral tonic-clonic seizure (P = 3.74 x 10^{-5}) and abnormal spatial working memory mouse phenotypes (P = 1.64 x 10^{-6}) (Supplementary Data 12). These two identified gene-sets included glutamate and GABAergic receptors, highlighting a plausible pathway by which these gene sets may potentially contribute to pain mechanisms (Supplementary Data 13, Supplementary Data 14). Other cell-type-specific associations included microglia and gut B cells with pain duration GWASs, with genes mapped to mouse phenotypes reflecting altered immune cell numbers and function indicating that these pathways may play a role in the temporal aspects of pain (Figure 3D).

Transcription-Wide Association Study

To determine the genetic relationship between changes in gene expression and chronic pain phenotypes, we performed a Transcriptome-Wide Association Study (TWAS) using eQTLs for eight brain cell-types (25) and the tibial nerve (25) to capture both the central and peripheral aspects involved in chronic pain. We tested the genetic association of imputed gene expression and our chronic pain phenotypes and identified 34 tibial nerve and 120 brain cell-type specific significantly associated genes of which 18 (e.g., *NPC1*, *BPTF*) and 29 (e.g., *BSN, HLA-DRB1*) were also identified through GWAS, respectively. Twelve genes (e.g., *MST1R, GMPPB*) were found in both TWASs as well as the GWAS analyses (Figure 4A, Supplementary Data 15, Supplementary Data 16).

To improve our understanding of the underlying mechanisms and pathways involved in chronic pain, we created a chronic pain gene-set by combining the genes from our chronic pain GWAS, rare variant burden associated genes, and TWAS, alongside 57 previously published pain genes (7) and performed Gene-Ontology (GO) analysis on this gene-set (Supplementary Data 16). GO analysis highlighted regulation of neuronal projection, endosomal pathways including postsynaptic recycling, and the plasma membrane (Figure 4B, Supplementary Data 17). Integration of the TWAS data with the GO findings can help in assigning directionality of pathway changes by providing information on gene expression changes within cell-types. Our results show that direction of effects can be cell-type specific--for example, *CERS2* was significantly associated in endothelial cells and the tibial nerve TWASs with observed effect in oppositive directions (Figure 4C).

ARTICLE IN PRESS

ARTICLE IN PRESS

Shared genetics between analgesic-use defined pain phenotypes and other pain-related diseases

We performed LD Score Regression (LDSC) analyses to investigate the degree of genetic heritability and shared genetic architecture between the 11 prescription-based pain phenotypes and 12 GWASs of conventional pain-related traits (fibromyalgia, migraine, and osteoarthritis). This analysis would shed light to the approach we applied to define pain phenotypes using prescription-based data to identify genes and pathways involved in pain. The sample sizes varied by condition, ranging from 85,708 for fibromyalgia to 1,174,790 for osteoarthritis (Supplementary Data 18). The proportion of variation in chronic pain explained by common genetic variants was similar across the 11 prescription-based pain phenotypes (0.002-0.04) (Supplementary Data 19) and to that estimated for migraine, fibromyalgia, and osteoarthritis (0.008 – 0.01) (Supplementary Data 18) highlighting that we are likely capturing true pain patients.

Focusing on shared genetic architecture, we found the strongest positive genetic correlations between fibromyalgia and our measures of chronic pain susceptibility, and pain severity (r_g ranging from 0.8 to 0.9, $P < 4.8 \times 10^{-5}$). These results imply that fibromyalgia is more closely related to pain sensitisation than the temporal aspects of pain given there was no evidence of significant genetic correlation between fibromyalgia and pain duration phenotypes ($P > 4.8 \times 10^{-5}$). Additionally, moderate significant positive genetic correlations ($r_g > 0.5$, $P < 4.8 \times 10^{-5}$) were observed between our chronic pain phenotypes and osteoarthritis and migraine (r_g 0.50 - 0.78, $P < 4.8 \times 10^{-5}$). To delineate whether these correlations were likely linked to sensitisation or time-related characteristics of pain, we compared the level of genetic correlation for pain duration to pain severity phenotypes and found moderate significant genetic correlations between pain duration and severity phenotypes with osteoarthritis, (r_g ranging from 0.52 to 0.70, $P < 4.8 \times 10^{-5}$) but only moderate significant genetic correlations between migraine and pain duration phenotypes (r_g ranging from 0.51 to 0.55, $P < 4.8 \times 10^{-5}$) (Figure 5, Supplementary Data 20). For sex-specific genetic effects, we observed strong significant genetic correlations between pain severity or pain duration and fibromyalgia in males ($r_g > |0.8|$, $P < 4.8 \times 10^{-5}$) but only moderate significant genetic correlation in females ($0.6 > r_g < |0.8|$, $P < 4.8 \times 10^{-5}$) (Supplementary Figure 4, Supplementary Data 20). Similar to the combined analyses, migraine showed moderate significant genetic correlation with pain duration as opposed to pain severity in sex-stratified analyses ($r_g > |0.5|$, $P < 4.8 \times 10^{-5}$; Supplementary Figure 4, Supplementary Data 20).

Shared genetic architecture of pain with neuropsychiatric and immunological traits.

Many studies have demonstrated a link between chronic pain and psychiatric conditions, mental health conditions and immune related disorders (9, 27-29). To determine if our chronic pain phenotypes share genetic components with these conditions, we performed LDSC on 21 traits spanning across psychiatric (e.g., bipolar, depression and schizophrenia), immune-related traits (e.g., Crohn's, Inflammatory Bowel Disease (IBD) and rheumatoid arthritis), and anthropometric traits (e.g., height & weight) (Supplementary Data 18). We selected these GWAS analyses due to their known comorbidity with pain related conditions, the availability of robust GWAS data, and their representation of diverse biological mechanisms. The strongest genetic correlations were found between pain duration phenotypes and depression or post-traumatic stress disorder (PTSD) ($r_g > |0.8|$, $P < 4.8 \times 10^{-5}$; Figure 5, Supplementary Data 20). Modest but consistent genetic correlations were also observed across the other pain categories (chronic pain, pain severity) with psychiatric conditions (e.g., anxiety and, PTSD) ($r_g > |0.5|$, $P < 4.8 \times 10^{-5}$; Figure 5, Supplementary Data 20). Despite other studies showing a link between chronic pain and immune-related disorders or anthropometric traits implying some shared mechanisms (9, 27-29), we did not find any evidence of significant genetic correlation between chronic pain and these

ARTICLE IN PRESS

phenotypes ($P < 4.8 \times 10^{-5}$; Figure 5, Supplementary Data 20). Chronic fatigue, a multifaceted disorder, was also included in LDSC analyses to determine the degree of genetic overlap with pain as the exact causes of chronic fatigue remain unknown (30). Consistent moderate genetic correlations were observed between chronic fatigue and our three pain phenotypes indicating that some shared genetic factors could influence pain symptoms and treatments given to chronic fatigue patients ($r_g > |0.5|$, $P < 4.8 \times 10^{-5}$; Figure 5, Supplementary Data 20).

Association of rare variants with chronic pain phenotypes

In addition to interrogating common variants associated with chronic pain phenotypes, we performed gene-based association testing of rare variants in UK Biobank (Supplementary Data 21). While we did not observe any exome-wide significant associations ($P < 1 \times 10^{-6}$) with any of our pain traits in UK Biobank, we did identify 66 genes with a suggestive association with chronic pain phenotypes ($P < 1 \times 10^{-4}$) (Supplementary Data 22). One of the rare variant gene-based suggestive associations (*PIK3C2A*) was also observed in our common variant GWAS of pain susceptibility ($P = 4.5 \times 10^{-10}$) (Supplementary Data 4, Supplementary Data 22). The variant-to-gene mapping at this locus was complex with QTL colocalization evidence for multiple genes. This additional genetic evidence highlighting a suggestive association between rare LoF variants within *PIK3C2A* and risk of chronic pain suggests that *PIK3C2A* is likely the candidate gene within the locus. We did not observe any rare variant gene-based suggestive associations ($P < 1 \times 10^{-4}$) with pain severity or pain duration phenotypes (Supplementary Data 22).

Discussion

We conducted a genome-wide association meta-analysis of chronic pain defined by analgesic prescription data in up to 692,039 individuals of European descent. We identified 140 GWAS associations for chronic pain phenotypes, of which 78 are novel. The number of identified loci is comparable to the number (125) identified by Toikumo *et al.* (11), who focused on pain intensity, and is higher than the 99 loci identified in prior chronic pain GWAS' (10, 14, 31-34). This approach of defining chronic pain patients via their medication usage offers a unique perspective on pain by removing the reliance upon self-perception and enabling identification of patients experiencing chronic pain who may never have received a specific pain diagnosis, resulting in the identification of additional pain association signals (e.g. *CNOT4, CNTNAP2*, *ARRP21*). Most GWAS loci were associated with susceptibility to chronic pain rather than pain severity or analgesic-use duration. Heritability across the pain phenotypes were similar suggesting that the fewer loci found for severity and duration may be due to a lack of statistical power compared to the larger susceptibility definitions.

The chronic pain locus we identified on chromosome 3 *p21.31* contains many plausible causal genes (Supplementary Figure 5). Although there is eQTL colocalization evidence for both *RBM6* and *CAMKV* (Supplementary Data 4), the strongest evidence points to *SEMA3F* for which there is coding variant evidence (Leu503Met, rs1046956:T:A, EAF = 0.72, OR = 0.96, $P = 1.4 \times 10^{-13}$) alongside eQTL colocalization evidence implicating that increased *SEMA3F* gene expression is correlated with increased chronic pain risk. SEMA3F is a class 3 semaphorin which are secreted ligands regulating axon guidance and ensuring appropriate pathfinding of sensory afferents in the spinal cord (35). Semaphorins bind to neuropilins and plexins, and there is growing evidence for the involvement of axon guidance molecules more generally in chronic pain disorders. For example, *DCC,* which encodes the netrin-1 receptor and regulates axon guidance, is repeatedly identified in genetic studies of pain, including our own (36) and Plexin-B1-RhoA signalling (with RhoA also identified in our TWAS - Supplementary Data 13) is implicated in inflammatory pain in mice (38), and in humans an auto-antibody against Plexin-D1 has been proposed

ARTICLE IN PRESS

ARTICLE IN PRESS

as pathogenic in neuropathic pain (39). Furthermore, these candidate genes from our GWAS and TWAS were significantly functionally associated with a common neuronal influence (GO term: Negative Regulation of Neuronal Projection Development, $P = 7.5 \times 10^{-4}$). Given the strong analgesic effects of therapies targeting the neurotrophic factor, NGF, which nevertheless has been limited by degenerative effects on the joint (40), the evaluation of safer therapeutic opportunities in axonal guidance molecules deserves further exploration.

Sex-stratified analyses revealed 21 female-specific associations and 11 male-specific associations indicating potential sex-specific effects. Similarly to the combined analyses, the majority were associated with susceptibility to chronic pain rather than severity of pain or duration of analgesic use. Sex-stratified loci were only associated with pain severity in males implicating that male pain perception may be more genetically driven than female pain perception. We found a robust association for *PER3* (Asp852His, rs145213510:G:C, OR = 2.02, $P=3.7 \times 10^{-8}$. EAF = 0.002) as a female-specific locus associated with chronic pain susceptibility (Supplementary Data 6, Supplementary Data 7, Supplementary Figure 6). *PER3* is a clock gene, which has not previously been associated with pain, but is a risk gene for morning chronotype and insomnia (41, 42); structural polymorphisms are associated with delayed sleep-phase syndrome (43) and mouse *PER3* knock-outs have shorter circadian periods (44). Both human pain sensory thresholds and chronic pain disorders (46) have strong diurnal patterns which may reflect direct involvement of circadian clock genes (47) or the influence of circadian-dependent processes such as sleep-wake cycles (48) on pain mechanisms. Sex-specific effects on circadian rhythm have been reported with women having shorter circadian periods (49). *PER3* may therefore affect pain susceptibility in women indirectly through changing the circadian period or affecting sleep (50) and improvements in circadian alignment through lifestyle changes or melatonin treatments could reduce susceptibility to chronic pain disorders. Further studies are required to determine how the *PER3* missense variant (Asp852His) affects protein function and confirm the likely role *PER3* has in pain.

Additionally, we also identified *OPRM1* as a male-specific locus associated with chronic pain susceptibility and duration with evidence of eQTL colocalization indicating higher *OPRM1* mRNA expression is associated with reduced time on analgesics (Supplementary Data 6, Supplementary Data 7, Supplementary Figure 7). *OPRM1* encodes the mu-opioid receptor, which is the primary site of action of endogenous opioid peptides (51). We previously showed using transcriptomic data from mouse models of pain, that suppression of the opioid signalling network including Oprm1, is evident in the transition from acute to chronic pain (52). Furthermore, polymorphisms in *OPRM1* are associated with differential response to opioids (53). The male-specific association in our study is interesting and aligns with previous studies showing differential efficacy of opioids in males and females (54, 55) . Genetic risk factors altering the efficacy of opioids could in turn lead to altered use behaviour and dependence. Better understanding of the role of *OPRM1* could improve chronic pain outcomes including developing approaches to reduce the risk of prolonged use of exogenous opioids.

Gene burden-test analysis, despite not identifying any statistically significant associations ($P < 1 \times 10^{-6}$), did identify 66 genes suggestively associated with chronic pain (at $P < 1 \times 10^{-4}$) (Supplementary Data 22). This is the largest such study to date and identifies genes not previously reported as associated with chronic pain. The overlap with loci identified through our chronic pain GWAS was limited with only one gene - *PIK3C2A* – being identified in both. Identifying an association between both *PIK3C2A* common and rare variants and pain susceptibility phenotypes increases confidence that *PIK3C2A* is likely to be the candidate gene for the GWAS association. *PIK3C2A* encodes an enzyme in the phosphoinositide 3-kinase (PI3-kinase) family which are involved in intracellular protein trafficking, cell survival and the inflammatory response implicating these pathways as potentially important in determining pain

susceptibility (56). Several other genes showing the strongest suggestive associations in gene burden-test analysis, despite not overlapping with GWAS loci, are involved in relevant biological mechanisms in pain. For example, *TNFAIP8L1*, otherwise known as TIPE1, is part of a family of immunity regulators, interacting with Rac1 to increase caspase-mediated apoptosis (57). Caspase-mediated apoptosis has previously found to play a critical role in neuropathic and inflammatory pain and could be a potential target for pharmacological interventions (58, 59). The burden of rare loss-of-function (LoF) variants within *IGF2BP2* were suggestively associated with chronic pain ($P = 4.4 \times 10^{-5}$; Supplementary Data 22) and another member of the same gene family, *IGF2BP3*, was prioritised through GWAS ($P = 1.5 \times 10^{-8}$; Supplementary Data 4). These insulin-like growth factor 2 mRNA binding proteins may affect the pain pathological process through regulation of RNA metabolism, transcriptional regulation of pain-related genes and release of inflammatory mediators (60-62). The identification of these rare variant associations merit further investigation as they may impart larger effect sizes than common variants and highlight important chronic pain pathophysiological mechanisms.

We show a high degree of genetic correlation between our chronic pain phenotypes and specific conventionally defined pain conditions, such as fibromyalgia, migraine, and osteoarthritis, with fibromyalgia demonstrating the strongest significant genetic correlations ($r_g > |0.8|$, $P < 4.8 \times 10^{-5}$). Fibromyalgia is a phenotype difficult to define, with a prevalence of up to 2.2% in our study cohort, and has limited well-powered published GWAS studies. This strong genetic correlation with chronic pain phenotypes indicates that the use of our prescription-based phenotyping can potentially identify more fibromyalgia patients than is formally diagnosed and can detect relevant pain related genes that would otherwise be missed using diagnostic criteria. Migraine showed modest significant genetic correlations (Bonferroni corrected $P < 4.8 \times 10^{-5}$) with chronic pain duration and susceptibility ($r_g > |0.5|$). Significant genetic correlations were also seen between migraine and pain severity phenotypes as well ($P < 4.8 \times 10^{-5}$) but the correlation coefficient was just below our threshold of 0.5 ($r_g = 0.48$). This weaker correlation coefficient may be driven by smaller sample sizes in the pain severity phenotypes. Osteoarthritis was found to be genetically correlated with chronic pain susceptibility, severity, and duration phenotypes consistent with osteoarthritis being a progressive chronic pain disorder ($r_g > |0.5|$, $P < 4.8 \times 10^{-5}$).

ARTICLE IN PRESS

Chronic pain is frequently co-morbid with mental health disorders, and we found evidence of shared genetics with psychiatric traits. These findings are consistent to those reported in previous GWAS' for other pain-related phenotypes (10, 11, 34). In contrast we did not identify genetic correlation of chronic pain with immune disorders, suggesting distinct underlying genetic mechanisms. This does not preclude shared environmental risk factors that may instead explain the frequent comorbidity of pain and immune disorders. Further interrogation of which genes and potential pathways are shared between chronic pain and these conditions will yield better pathophysiological understanding, improved clinical care and could lead to development of therapies to effectively treat both.

The expression and functional enrichment analyses highlight potential underlying mechanisms involved in chronic pain. Using cell-type specific enrichments and mouse phenotype gene-sets, we found association of susceptibility to chronic pain requiring opiates with oligodendrocyte progenitor cells (OPCs), gut neurons and peptidergic nociceptors. Chronic pain duration, defined by extended analgesic use, was associated with neuronal and inflammatory cell-types, namely microglia and gut B cells. Interestingly migraine GWAS loci (16) were also enriched in microglia and was correlated with the analgesic use duration phenotype, highlighting modulation of microglia as a therapeutic opportunity in migraine.

ARTICLE IN PRESS

Given the lack of other cohorts with sufficient linkage of genetic data with prescription data, additional studies will be required to validate our approach. It is however reassuring that our study re-identified many previously associated chronic pain loci as well as potentially novel loci. Utilisation of a UK Biobank reference panel in LDSC analysis may have introduced biases in the estimation of LD scores given that our study also included Finnish individuals (63, 64). Specifically, genetic variation, allele frequencies, and LD patterns can differ significantly between populations of varying ancestries (65). These differences can lead to inaccuracies in LDSC and impact the reliability and interpretation of heritability estimates and genetic correlations derived from the analysis(66-68). Additionally, the participants from the United Kingdom and Finland included in our cohorts have accessible healthcare systems and are generally healthier individuals with a higher socioeconomic status (66-68). Caution is, therefore, needed when generalising our findings and our phenotypic strategy of using electronic prescription data to accurately identify chronic pain cases and controls beyond the ethnic populations and specific country healthcare systems represented in this study. Sensitivity analyses using ethnically diverse patient cohorts and cohorts where interaction with healthcare systems varies will help validate genetic associations and strengthen conclusions from this study. The SNP-based heritability estimates for our chronic pain phenotypes are slightly lower than that estimated for other single-pain conditions using the same data. These low SNP-based heritability estimates may be attributable to chronic pain being a multifactorial condition with many determinants, may be due to our cohorts comprising of relatively healthy individuals who were not specifically recruited for chronic pain or may be because prescription drug use can be influenced by many factors. Heritability estimates derived using family studies and using patients specifically recruited for pain could provide more accurate values (69). A further limitation to our approach is that chronic pain patients treated for less than 3 months or not treated with an analgesic and instead provided with a non-analgesic pain treatment will not be captured by our study. This is most likely to impact the ascertainment of neuropathic pain conditions.

Collectively, our advances in pain phenotyping paired with the integration of GWAS, rare-variant gene-based tests, and TWAS highlights genes implicated in pain that may have high public significance given current unmet medical need for chronic pain and challenges with existing analgesic therapies. Pathway enrichment highlights neuronal guidance pathways (containing genes including *DCC*, *SEMA3B* and *SEMA3F*), the endolysosomal and post-synaptic recycling endosomal pathways. This should help direct future cellular and *in vivo* studies to elucidate pain mechanisms. A similar prescription-based phenotyping approach could be taken for other heterogeneous traits which are also poorly diagnosed in the clinic, leading to further understanding of the underlying pathophysiology of disease and eventual potential identification of therapeutic targets.

Methods

Ethics

The UK Biobank study was approved by the North-West Multi-Centre Research Ethics Committee (ref. 11/NW/0382 on June 17, 2011). All participants provided written informed consent to participate in the UK Biobank study. The study protocol is available online (http://www.ukbiobank.ac.uk/). This research has been conducted using the UK Biobank resource under application number 20361.
Study subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea

ARTICLE IN PRESS

ARTICLE IN PRESS

(Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017.

Overview of analyses

An overview of the analyses is outlined in Figure 6. We defined 11 chronic pain phenotypes (Supplementary Data 1) using prescription data from electronic health records in UK Biobank and FinnGen (Supplementary Data 2). We then performed GWAS followed by meta-analysis of each of the 11 phenotypes alongside sex-stratified analyses in European participants. We also performed gene-based tests using UK Biobank whole genome sequencing (WGS) data. Downstream analysis included prioritisation of candidate genes for the GWAS associations, estimated heritability, genetic correlations, alongside functional enrichment analyses including pathways, mouse phenotypes and cell-types.

UK Biobank

The UK Biobank is a large prospective population-based population with over 500,000 participants aged 40-69 recruited across the UK between 2006-2010 (70). UK Biobank contains extensive phenotypic and genotypic details about participants including questionnaire data, physical measures, samples, accelerometery, genome-wide genotyping and longitudinal follow-up for a range of health-related outcomes. Primary care data, providing structured longitudinal and prescription data, for ~230,000 participants was made available in 2019 and was used for phenotype definition. UK Biobank obtained informed consent from all participants. Given that the incomplete linkage of primary care data in the UK Biobank, we required any control populations to only be selected from subset of ~230,000 UKB participants with primary care linkage. Whilst UK Biobank includes a sizeable minority with non-European ancestries, genetic analysis on this subgroup was not undertaken as the primary goal was to complete a meta-analysis with FinnGen.

ARTICLE IN PRESS

FinnGen

The FinnGen study is a large-scale genomics initiative that has analyzed over 500,000 Finnish biobank samples and correlated genetic variation with health data to understand disease mechanisms and predispositions (67). The project is a collaboration between research organisations and biobanks within Finland and international industry partners. FinnGen provides nationwide longitudinal digital health record data from Finnish health registries collected since 1969. The registries contain coverage on all health-related events, including hospitalizations, prescription drug purchases, medical procedures or deaths. Participants provided informed consent for biobank research.

Phenotype Definitions

Example code for producing the phenotypes is provided in the Supplementary Software 1. We constructed 11 binary phenotypes describing chronic pain, pain severity, and pain duration (Supplementary Data 1). These phenotypes cover both comparisons to a ‘healthy control’ population (e.g. individuals that did not meet our core definition of chronic pain); and, comparisons within the chronic pain population.

A clinician reviewed a comprehensive list of analgesics prescribed within both UK Biobank and FinnGen. Analgesics were additionally annotated as weak analgesics (e.g. paracetamol, non-steroidal anti-inflammatory drugs (NSAIDs), selective COX inhibitors, triptans and weak opioids) or strong opioids (e.g. buprenorphine, diamorphine, fentanyl, hydromorphone, meptazinol, methadone, morphine, oxycodone,

ARTICLE IN PRESS

pethidine, tapentadol & tramadol). We assumed that each analgesic prescription is for 30 days duration. While it is possible that actual prescription medication use may not encompass the full 30-day period, our study focused on the intervals between prescriptions to identify both uninterrupted analgesic use and to identify cessation of treatment rather than focusing on the duration of a particular prescription.

The 'Chronic Pain' case population comprised of individuals aged ≥18 years with evidence of chronic pain based on at least one period of analgesic prescriptions ≥90 days in duration. A 'period of analgesic prescriptions' was defined when: 1) the interval between the date of first analgesic prescription and the date of the last analgesic prescription is ≥90 days duration; and, 2) there is less than a 45 day break between the start of each prescription (assuming a 30-day prescription duration +/- 15 day buffer between prescriptions). The index date for chronic pain was defined as 90 days after the date of first analgesic prescription.

Total analgesic duration ('Time on Analgesics') for each patient was quantified by summing the duration of each analgesic prescribing period. Individuals were considered to have ceased analgesic treatment (i.e. entered remission or successfully adopted an analgesic free treatment approach) when a break in analgesic prescription issue dates exceeded 180 days. We defined the time-to-cessation ('Cessation Time') for each patient as the number of days from the chronic pain index date to the last analgesic issue date prior to commencement of a period of >180 days without an analgesic prescription. For both 'Time on Analgesics' and 'Cessation Time', we divided the chronic pain population into quartiles. Our phenotypes then compared: the highest quartile (Q4) to the lowest quartile (Q1) within the chronic pain population; the highest quartile (Q4) to all other quartiles (Q1-Q3) within the chronic pain population; and, the highest quartile (Q4) to healthy controls.

ARTICLE IN PRESS

Genotyping and Imputation

UK Biobank

Genotyping, quality control, and imputation was performed centrally by UK Biobank as described by Bycroft *et al.* (71) with a total of 97 million variants. We filtered the genotype data to only include variants with minor allele count ≥ 25, imputation quality score (INFO) ≥ 0.3, and Hardy-Weinberg equilibrium exact test *P* value > 1×10^{-12}. We required a maximum per-variant and per-sample missing call rate <0.01. The final target dataset included 42 million single nucleotide polymorphisms (SNPs) after excluding all duplicate SNPs. Only individuals of European ancestry as defined by the Pan UKBB (https://pan.ukbb.broadinstitute.org/) were included in analyses. The Pan UKBB ancestry assignments are available for download through the UK Biobank portal as Return 2442 (https://biobank.ctsu.ox.ac.uk/showcase/dset.cgi?id=2442). Details about the QC process from the Pan UKBB can be found here including determination of ancestry groups (https://pan.ukbb.broadinstitute.org/docs/qc/index.html)). Genetic sex was defined according to UK Biobank Field 22001 (https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=22001).

FinnGen

We used genotype data from FinnGen Release 11 (R11). FinnGen individuals were genotyped with Illumina and Affymetrix chip arrays (Illumina Inc., San Diego, and Thermo Fisher Scientific, Santa Clara, CA, USA). Chip genotype data were imputed using the population-specific SISu v4.2 imputation reference panel of 8,554 whole genomes. Merged imputed genotype data is composed of 116 data sets that include samples from multiple cohorts. In sample-wise quality control steps, individuals with sex-reported gender mismatch, high genotype missingness (>5%), excess heterozygosity (+-4SD) and non-Finnish ancestry were excluded. In variant-wise quality control steps, variants with high missingness

ARTICLE IN PRESS

(>2%), low HWE *P* value (<1 x 10^{-6}) and low minor allele count (MAC < 3) were excluded. The final dataset included 21 million SNP for *N*=473,681 individuals.

Association analyses and risk loci definition

GWAS analyses were performed on each of the 11 phenotypes in both UK Biobank and FinnGen adjusting for age at chronic pain (for analyses within the chronic pain population) or age at baseline assessment (for analyses comparing chronic pain to healthy controls), genetic sex, genotype chip, and the first 20 principal components. Logistic regression using an additive genetic model was run in the Regenie software (72). We then performed an inverse-variance weighted fixed effects meta-analyses of the 21 million SNPs available in both biobank on each of the 11 phenotypes.

To define independent risk loci, the variant with the lowest *P* value across the genome was defined as an index variant and a 1 mega base (Mb) locus centered at the index variant was defined. This was repeated until no further variants reaching genome-wide significance (*P* values < 5 x 10^{-8}) remained.

To assess the overlap of genetic associations across the 11 pain phenotypes, unique loci were identified using an arbitrary 1 Mb distance of one index variant from another index variant. The VennDiagram (73) and the UpSetR (74) packages were used in R (75) to determine the overlap of unique loci across pain phenotypes (chronic pain, pain duration and pain severity) and across sexes.

To determine whether genetic associations identified in our meta-analyses of our 11 chronic pain phenotypes had been previously reported, we compared our genetic findings to those recently reported in the literature for pain or pain-related conditions, including pain susceptibility (10), general pain factor (33), pain intensity (11) and previously reported pain genes reviewed by Li *et al.* (7) (Supplementary Data 20). We also compared to previously reported loci associated with two additional pain-related disorders: osteoarthritis (15) and migraine (16) (Supplementary Data 20). To align with the human assembly used in this study (GrCh38), the chromosome and position of the loci from each study were converted from GrCh37 to GrCh38 using UCSC's LiftOver tool (76).

Rare variant burden testing

Rare-variant gene-based burden analyses were performed using WGS data from the UK Biobank. The WGS methods are described in detail in Li *et al.* (77). Briefly, 490,640 UK Biobank participants were sequenced to an average depth of 32.5X using the Illumina NovaSeq 6000 platform. Variants were jointly called using Graphtyper, which resulted in 1,037,556,156 and 101,188,713 high quality (AAscore < 0.5 and < 5 duplicate inconsistencies) SNPs and indels respectively. We further processed the jointly called genotype data in Hail v0.2, where multi-allelic sites were first split and normalized. Variants were then filtered based on low allelic balance (ABHet < 0.175, ABHom < 0.9), low quality-by-depth normalized score (QD < 6), low phred-scaled quality score (QUAL < 10) and high missingness (call rate < 90%).
We defined a cohort of European ancestry individuals to be that most resembled the NFE (non-Finnish European) population as labelled in the gnomAD v3.1 dataset (78). Variant loadings for 76,399 high-quality ancestry-informative variants from gnomAD were used to project the first 16 principal components onto all UK Biobank WGS samples. A random forest classifier trained on the nine gnomAD ancestry labels was then used to assign individuals to ancestry groups

ARTICLE IN PRESS

(probability > 0.9). In total, 458,855 individuals of European ancestry were taken forward for analysis.
Variants were annotated using Ensembl's VEP tool (79) and putative loss of function (LoF) variants were annotated using the LoFTEE plugin.
For burden testing, we defined five masks using variants with minor allele frequency (MAF) < 0.001: 1) LoFHC (LoFTEE high confidence), 2) All LoF (LoFTEE high and low confidence), 3) All LoF and DeleteriousMissense (combination of mask 2 and SIFT or Polyphen prediction being deleterious or probably damaging respectively), and 4) All LoF & Missense (combination of mask 2 and any variant annotated as a missense by VEP), and 5) All Missense (any VEP annotated missense variant). For each phenotype, gene-based burden testing was performed using the SKATO method as implemented in Regenie v3.2.5 (72) on the same phenotypes as the UK Biobank GWAS.

SNP-based heritability and Genetic correlations

LD score regression (LDSC) (80) was used to estimate per-trait SNP-based heritability (h^2) and pairwise genetic correlation (r_g) between the our 11 chronic pain phenotypes and GWAS meta-analyses that focused on pain traits (migraine and osteoarthritis), psychiatric conditions (bipolar disorder, depression, anxiety, post-traumatic stress disorder (PTSD)), and immune-related conditions (rheumatoid arthritis, Crohn's disease, inflammatory bowel disease (IBD)) (Supplementary Data 23).

LD scores used in the analyses were derived using LDSC from 10000 unrelated non-Finnish European individuals from UK Biobank with available whole genome sequencing (WGS) data. LD scores were estimated by collapsing common (MAF > 0.01) HapMap3 variants within 1Mb windows to comprehensively capture common variant tagging. Variants included in the analyses were restricted to well-imputed HapMap3 variants, per method recommendations.

For some traits, particularly those which were sex-stratified, low statistical power led to inaccurate estimations of SNP-based heritability ($h^2 \sim 0.01$ or $h^2 < 0$). This meant that pairwise genetic correlation could not be estimated for a subset of trait combinations. The statistical significant threshold was set at 4.8×10^{-5} (0.05 / 1087) where 1087 is the number of independent tests.

Prioritisation of most likely causal genes

To determine the most likely causal gene at each loci, we performed three complementary approaches: 1. 'Nearest Gene': we identified the nearest protein-coding gene to the lead variant based on distance to the transcription start site or the gene body; whichever was closer. 2. 'Coding variant': we used VEP (79) to annotate variants with their predicted effect and considered any coding variant with high or moderate impact (missense, frameshift, stop gained, start/stop lost, inframe insertion/deletion) within the credible set of the signal with a posterior probability > 0.01 (81) as evidence for prioritisation. 3. 'Quantitative trait locus (QTL) colocalization': We performed colocalization of the pain GWAS signals with pQTLs and eQTLs. For eQTLs, we used data from the Genotype-Tissue Expression Consortium (https://gtexportal.org, (25)), BLUEPRINT (96) and MetaBrain (97). For pQTLs, we used data from UK Biobank (82), DeCODE (83) and Fenland (84). For both eQTLs and pQTLs, we considered only those acting in *cis,* defined as genes located < 1 MB from the lead variant. We performed multicolocalization analysis using the *coloc* package (85) in R (75). We determined a significant colocalization if the GWAS signal had a $P < 5 \times 10^{-8}$, the QTL signal had a $P < 1 \times 10^{-5}$ and the posterior probability for the signals being shared was greater than 0.8 ($H12 > 0.8$).

ARTICLE IN PRESS

Transcriptomic Wide Association Analysis

As chronic pain involves both the central and peripheral nervous system, we focused transcriptome-wide association (TWAS) analyses on brain cell-types (central nervous system) and tibial nerve (peripheral nervous system). Using the S-PrediXcan software (86), we analysed cis-eQTLs from Bryois *et al.* (87) for eight brain cell-types derived from 373 human brain samples from 215 individuals, as well as tibial nerve eQTLs from the GTEx project (25). We tested the genetic association between chronic pain phenotypes and both cell-type-specific and tibial nerve eQTLs. We identified TWAS significant genes as those that passed mfxcultiple testing corrections for the number of genes analysed in the TWAS for each tissue and cell-type. The TWAS process involves two stages. First, gene expression and genotype datasets are integrated using a regression model, assuming linear additive genetic effects. In the second step, the eQTL effect sizes are used to impute gene expression in an independent GWAS dataset, and the association between the trait and imputed gene expression is tested.

Gene-set cell-type enrichment

We used publicly available datasets from published studies, including the cortex, substantia nigra (20), trigeminal ganglia (22), dorsal root ganglia (22), and gut (22, 23). The cell-types in each single-cell dataset are listed in Supplementary Data 24. The available information for transcriptomic data sets were in two formats: (1) the mean expression for each gene in each cell type and (2) the gene expressions count matrix. If the mean expression for each gene is available, we used to identify specifically expressed genes and the 10% most expressed genes for each cell type are defined as those with the large mean expression. However, if the gene expression matrix is available, we computed t-statistics for each gene to identify the most specifically expressed genes. To compute a t-statistic, each gene (*g*) was assigned a value of "1" if the sample is expressed in a cell-type and "-1" if otherwise. We define outcome (*y*) as scaled gene expression and then fit a simple linear regression model of "*y*" on "*g*" and compute t-statistics for each gene. This will produce t-statistics for each gene in each cell-type, and the 10% most expressed genes for a given cell-type are defined as those with the largest positive t-statistics.

We used MAGMA (88) (v1.10) software to perform gene-set enrichment analysis using GWAS summary statistics datasets. Using the GWAS summary statistics dataset, MAGMA performs gene-set enrichment in two steps, (1) gene-level association and (2) gene-set enrichment. In the gene-level association, MAGMA averages SNPs *P* values located around the gene while accounting for LD. In the gene-set analysis, the averaged gene-level *P* values are converted to z-scores and used to test whether a gene-set is more enriched in a phenotype compared to genes not in the gene-set. A 25 kilobases (kb) window upstream and downstream on a gene location was used for gene-level analysis to define a gene window. MAGMA was then used to investigate whether the 10% most specifically expressed gene in each cell-type was associated with the phenotypes. A 5% significant threshold was set to determine significant association after Bonferroni correction.

Partitioned LD score (89) regression (LDSC) was used to investigate whether the 10% most specifically expressed genes in a cell-type account for the heritability of the phenotypes. We used a 25 kb window on both sides of a gene coordinate to define a gene region. The Bonferroni corrected *P* values were used to determine the association of a cell-type to a

phenotype and a 5% multiple testing cut-off was used as a threshold. The abstract mediation model (AMM) (90) was also used to test whether the 10% most expressed genes of each cell-type mediated the heritability in each trait. AMM estimates the fraction of heritability mediated by the kth-closest gene to each SNP because in the genome some fraction of heritability is mediated by genes that are not the closest. To account for gene contribution, the 50th-closest gene was used as the farthest gene from a SNP. Bonferroni corrected *P* values were used to determine the cell-types enriched for the heritability of the phenotypes.

We performed gene-set cell-type enrichment on our 11 chronic pain phenotypes defined using analgesic data alongside migraine (16). We included migraine to enable comparison of our results to a chronic pain disorder with large GWAS summary statistics available.

Gene-set functional characterisation

To characterise mouse phenotypes associated with pain-related cell-types, the top 10% most specifically expressed genes in each cell-type were tested for enrichment in mouse phenotype gene-sets. Mouse genotype-phenotype associations were obtained from the Monarch Initiative (24), filtered for genotypes affecting a single mouse gene with a high-confidence human ortholog as defined by Ensembl Compara (91), and then aggregated to the gene level. Enrichment was tested using the hypergeometric test implemented with the clusterProfiler R package (92) and was relative to the background set of genes detected in each scRNA-seq dataset. Significantly enriched gene-sets were defined as those with >= 2 overlapping genes, gene ratio >= 0.01, and FDR-adjusted *P* value <= 0.05. Cell-types shown to have cell-type-specific genes significantly enriched in pain-related GWAS signals were selected for further study. This included extracting cell-type-specific genes in enriched mouse phenotype gene-sets to test for pain GWAS enrichment.

ARTICLE IN PRESS

Pathway enrichment analysis and protein-protein interactions

We constructed a protein-protein interaction (PPI) network to represent the structure of interactions in cell-types associated with specific phenotypes. To do this, the PPI for the list of genes in each cell-type were selected and prepared for network analysis using the *igraph* (93) package in R (75). The *cluster_louvain* function was used for building the interactions. This clustering generated modules for each cell-type and modules with at least 30 genes were considered as more informative. The structure of the networks was further simplified to get useful information from them by using only genes with at least 10 interactions. Using *topGO* (94), and *rrvgo* (95) packages in R (75) we perform gene ontology enrichment analysis.

Data Availability

The meta-analysis summary statistics generated in this study are provided in the GWAS catalogue (https://www.ebi.ac.uk/gwas/) under the following accession codes: GCST90705000 [https://www.ebi.ac.uk/gwas/studies/GCST90705000], GCST90705001 [https://www.ebi.ac.uk/gwas/studies/GCST90705001], GCST90705002 [https://www.ebi.ac.uk/gwas/studies/GCST90705002], GCST90705003 [https://www.ebi.ac.uk/gwas/studies/GCST90705003], GCST90705004 [https://www.ebi.ac.uk/gwas/studies/GCST90705004], GCST90705005 [https://www.ebi.ac.uk/gwas/studies/GCST90705005], GCST90705006 [https://www.ebi.ac.uk/gwas/studies/GCST90705006], GCST90705007 [https://www.ebi.ac.uk/gwas/studies/GCST90705007], GCST90705008 [https://www.ebi.ac.uk/gwas/studies/GCST90705008], GCST90705009

ARTICLE IN PRESS

[https://www.ebi.ac.uk/gwas/studies/GCST90705009], GCST90705010
[https://www.ebi.ac.uk/gwas/studies/GCST90705010], GCST90705011
[https://www.ebi.ac.uk/gwas/studies/GCST90705011], GCST90705012
[https://www.ebi.ac.uk/gwas/studies/GCST90705012], GCST90705013
[https://www.ebi.ac.uk/gwas/studies/GCST90705013], GCST90705014
[https://www.ebi.ac.uk/gwas/studies/GCST90705014], GCST90705015
[https://www.ebi.ac.uk/gwas/studies/GCST90705015], GCST90705016
[https://www.ebi.ac.uk/gwas/studies/GCST90705016], GCST90705017
[https://www.ebi.ac.uk/gwas/studies/GCST90705017], GCST90705018
[https://www.ebi.ac.uk/gwas/studies/GCST90705018], GCST90705019
[https://www.ebi.ac.uk/gwas/studies/GCST90705019], GCST90705020
[https://www.ebi.ac.uk/gwas/studies/GCST90705020], GCST90705021
[https://www.ebi.ac.uk/gwas/studies/GCST90705021], GCST90705022
[https://www.ebi.ac.uk/gwas/studies/GCST90705022], GCST90705023
[https://www.ebi.ac.uk/gwas/studies/GCST90705023], GCST90705024
[https://www.ebi.ac.uk/gwas/studies/GCST90705024], GCST90705025
[https://www.ebi.ac.uk/gwas/studies/GCST90705025], GCST90705026
[https://www.ebi.ac.uk/gwas/studies/GCST90705026], GCST90705027
[https://www.ebi.ac.uk/gwas/studies/GCST90705027], GCST90705028
[https://www.ebi.ac.uk/gwas/studies/GCST90705028], GCST90705029
[https://www.ebi.ac.uk/gwas/studies/GCST90705029], GCST90705030
[https://www.ebi.ac.uk/gwas/studies/GCST90705030], GCST90705031
[https://www.ebi.ac.uk/gwas/studies/GCST90705031], GCST90705032
[https://www.ebi.ac.uk/gwas/studies/GCST90705032]). The full eQTL summary statistics used to prioritise most likely causal genes described in this manuscript were obtained directly from the GTEx Portal [https://gtexportal.org/home/], version 8, from BLUEPRINT[96] [http://blueprint-dev.bioinfo.cnio.es/WP10/qtls] and from https://www.metabrain.nl[97]. We downloaded and used all version 8 summary statistics across all available tissues from GTEx (downloadable from https://console.cloud.google.com/storage/browser/gtex-resources). The eQTL from BLUEPRINT are available on the European Genome-Phenome archive under accession number EGAD00001005199 [https://ega-archive.org/datasets/EGAD00001005199]. A request for access to the full summary statistics from metabrain was made via https://www.metabrain.nl/cis-eqtls.html. For the TWAS, the tibial nerve eQTL data was obtained from the GTEx Portal [https://gtexportal.org/home/], version 8. The full eQTL summary statistics for the eight brain specific cell-types (Inhibitory neurons, Pericytes, Oligodendrocytes, Oligodendrocyte Precursor Cells, Endothelial, Astrocytes, Excitatory neurons, Microglia) are available for download at https://doi.org/10.5281/zenodo.5543734. The full summary statistics for the pQTL data used to prioritise most likely causal genes are available for download from The full summary statistics for the pQTL data used to prioritise most likely causal genes are available for download from Zenodo [https://doi.org/10.5281/zenodo.2615265] and Omiscience [https://www.omicscience.org/apps/pgwas/]. The mean expression for each gene and the gene expression count matrix in the dorsal root ganglia are available for download at https://sensoryomics.shinyapps.io/RNA-Data/. The mean expression for each gene and the gene expression count matrix in the dorsal root ganglia are available for download at tg.painseq.com as well as within the Gene Expression Omnibus (GEO) repository [www.ncbi.nlm.nih.gov/geo] under accession number GSE197289 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE197289]. The mean expression for each gene and the gene expression count matrix in the substantia nigra are available from the Gene Expression Omnibus under the accession code GSE140231 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140231]. The summary statistics for the

ARTICLE IN PRESS

eQTLs for the eight brain cell-types used in the transcriptomic-wide association analyses are available for download on Zenodo [https://doi.org/10.5281/zenodo.5543734].The mean expression for each gene and the gene expression count matrix in the gut are available for download at https://www.gutcellatlas.org. Mouse genotype-phenotype associations were obtained from the Monarch Initiative [https://monarchinitiative.org] on https://data.monarchinitiative.org/dipper-kg/final/tsv/genotype_associations/genotype_phenotype.10090.tsv.gz and https://data.monarchinitiative.org/dipper-kg/final/tsv/all_associations/genotype_gene.all.tsv.gz. Human-mouse gene orthologs are available from Ensembl BioMart [https://www.ensembl.org/biomart/martview] under the Mouse genes dataset (GRCm39). Genome-wide summary statistics for migraine excluding 23andMe[16] are available for bona fide researchers by contacting Dale Nyholt (d.nyholt@qut.edu.au); see https://pmc.ncbi.nlm.nih.gov/articles/PMC8837554/#notes3. UK Biobank data was accessed under Application #20361. Instructions for accessing UK Biobank data are available at https://ams.ukbiobank.ac.uk/ams/. FinnGen individual-level data used for GWAS was accessed through GSK's membership in the FinnGen Consortia and the University of Helsinki. Documentation on accessing the FinnGen data is available at https://www.finngen.fi/en/researchers/accessing. Individual level data can only be accessed through the individual Biobanks themselves.
Summary statistics used to produce the regional association plots in the Response to Reviewer document are available for download at the GWAS catalogue [https://www.ebi.ac.uk/gwas/] under accession numbers GCST90705000 [https://www.ebi.ac.uk/gwas/studies/GCST90705000] & GCST90705015 [https://www.ebi.ac.uk/gwas/studies/GCST90705015]. The authors confirm that the remaining data supporting the findings of this study are available within the article and its supplementary materials

Code Availability

The source code used to produce the phenotypes used to generate all results in this publication are available in the Supplementary Software 1 file.

ARTICLE IN PRESS

References

1. Disease GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1211-59.

2. Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. Chronic pain as a symptom or a disease: the IASP Classification of Chronic Pain for the International Classification of Diseases (ICD-11). Pain. 2019;160(1):19-27.

3. Fayaz A, Croft P, Langford RM, Donaldson LJ, Jones GT. Prevalence of chronic pain in the UK: a systematic review and meta-analysis of population studies. BMJ Open. 2016;6(6):e010364.

4. Benyamin R, Trescot AM, Datta S, Buenaventura R, Adlaka R, Sehgal N, et al. Opioid complications and side effects. Pain Physician. 2008;11(2 Suppl):S105-20.

5. Camacho EM, Penner LS, Taylor A, Guthrie B, Avery AJ, Ashcroft DM, et al. Estimating the economic effect of harm associated with high risk prescribing of oral non-steroidal anti-

inflammatory drugs in England: population based cohort and economic modelling study. BMJ. 2024;386:e077880.
6. Noori A, Sadeghirad B, Wang L, Siemieniuk RAC, Shokoohi M, Kum E, et al. Comparative benefits and harms of individual opioids for chronic non-cancer pain: a systematic review and network meta-analysis of randomised trials. Br J Anaesth. 2022;129(3):394-406.
7. Li S, Brimmers A, van Boekel RLM, Vissers KCP, Coenen MJH. A systematic review of genome-wide association studies for pain, nociception, neuropathy, and pain treatment responses. Pain. 2023;164(9):1891-911.
8. Johnston KJA, Signer R, Huckins LM. Chronic overlapping pain conditions and nociplastic pain. HGG Adv. 2025;6(1):100381.
9. Johnston KJA, Adams MJ, Nicholl BI, Ward J, Strawbridge RJ, McIntosh AM, et al. Identification of novel common variants associated with chronic pain using conditional false discovery rate analysis with major depressive disorder and assessment of pleiotropic effects of LRFN5. Transl Psychiatry. 2019;9(1):310.
10. Mocci E, Ward K, Perry JA, Starkweather A, Stone LS, Schabrun SM, et al. Genome wide association joint analysis reveals 99 risk loci for pain susceptibility and pleiotropic relationships with psychiatric, metabolic, and immunological traits. PLoS Genet. 2023;19(10):e1010977.
11. Toikumo S, Vickers-Smith R, Jinwala Z, Xu H, Saini D, Hartwell EE, et al. A multi-ancestry genetic study of pain intensity in 598,339 veterans. Nat Med. 2024;30(4):1075-84.
12. Tsepilov YA, Freidin MB, Shadrina AS, Sharapov SZ, Elgaeva EE, Zundert Jv, et al. Analysis of genetically independent phenotypes identifies shared genetic factors associated with chronic musculoskeletal pain conditions. Communications Biology. 2020;3(1):329.
13. Song W, Lam M, Liu R, Simona A, Weiner SG, Urman RD, et al. A genome-wide Association study of the Count of Codeine prescriptions. Sci Rep. 2024;14(1):22780.
14. Zorina-Lichtenwalter K, Parisien M, Diatchenko L. Genetic studies of human neuropathic pain conditions: a review. Pain. 2018;159(3):583-94.
15. Boer CG, Hatzikotoulas K, Southam L, Stefansdottir L, Zhang Y, Coutinho de Almeida R, et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell. 2021;184(18):4784-818 e17.
16. Hautakangas H, Winsvold BS, Ruotsalainen SE, Bjornsdottir G, Harder AVE, Kogelman LJA, et al. Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles. Nat Genet. 2022;54(2):152-60.
17. Deak JD, Zhou H, Galimberti M, Levey DF, Wendt FR, Sanchez-Roige S, et al. Genome-wide association study in individuals of European and African ancestry and multi-trait analysis of opioid use disorder identifies 19 independent genome-wide significant risk loci. Mol Psychiatry. 2022;27(10):3970-9.
18. Kember RL, Vickers-Smith R, Xu H, Toikumo S, Niarchou M, Zhou H, et al. Cross-ancestry meta-analysis of opioid use disorder uncovers novel loci with predominant effects in brain regions associated with addiction. Nat Neurosci. 2022;25(10):1279-87.
19. Bartley EJ, Fillingim RB. Sex differences in pain: a brief review of clinical and experimental findings. Br J Anaesth. 2013;111(1):52-8.
20. Agarwal D, Sandor C, Volpato V, Caffrey TM, Monzon-Sandoval J, Bowden R, et al. A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. Nat Commun. 2020;11(1):4183.

ARTICLE IN PRESS

ARTICLE IN PRESS

21. Yang L, Xu M, Bhuiyan SA, Li J, Zhao J, Cohrs RJ, et al. Human and mouse trigeminal ganglia cell atlas implicates multiple cell types in migraine. Neuron. 2022;110(11):1806-21 e8.
22. Tavares-Ferreira D, Shiers S, Ray PR, Wangzhou A, Jeevakumar V, Sankaranarayanan I, et al. Spatial transcriptomics of dorsal root ganglia identifies molecular signatures of human nociceptors. Sci Transl Med. 2022;14(632):eabj8186.
23. Elmentaite R, Kumasaka N, Roberts K, Fleming A, Dann E, King HW, et al. Cells of the human intestinal tract mapped across space and time. Nature. 2021;597(7875):250-5.
24. Mungall CJ, McMurry JA, Kohler S, Balhoff JP, Borromeo C, Brush M, et al. The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res. 2017;45(D1):D712-D22.
25. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318-30.
26. Themistocleous AC, Baskozos G, Blesneac I, Comini M, Megy K, Chong S, et al. Investigating genotype-phenotype relationship of extreme neuropathic pain disorders in a UK national cohort. Brain Commun. 2023;5(2):fcad037.
27. Farrell SF, Kho PF, Lundberg M, Campos AI, Renteria ME, de Zoete RMJ, et al. A Shared Genetic Signature for Common Chronic Pain Conditions and its Impact on Biopsychosocial Traits. J Pain. 2023;24(3):369-86.
28. Meng W, Adams MJ, Reel P, Rajendrakumar A, Huang Y, Deary IJ, et al. Genetic correlations between pain phenotypes and depression and neuroticism. Eur J Hum Genet. 2020;28(3):358-66.
29. Gureje O, Von Korff M, Kola L, Demyttenaere K, He Y, Posada-Villa J, et al. The relation between multiple pains and mental disorders: results from the World Mental Health Surveys. Pain. 2008;135(1-2):82-91.
30. Das S, Taylor K, Kozubek J, Sardell J, Gardner S. Genetic risk factors for ME/CFS identified using combinatorial analysis. J Transl Med. 2022;20(1):598.
31. Suri P, Palmer MR, Tsepilov YA, Freidin MB, Boer CG, Yau MS, et al. Genome-wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. PLoS Genet. 2018;14(9):e1007601.
32. Johnston KJA, Ward J, Ray PR, Adams MJ, McIntosh AM, Smith BH, et al. Sex-stratified genome-wide association study of multisite chronic pain in UK Biobank. PLoS Genet. 2021;17(4):e1009428.
33. Zorina-Lichtenwalter K, Bango CI, Van Oudenhove L, Ceko M, Lindquist MA, Grotzinger AD, et al. Genetic risk shared across 24 chronic pain conditions: identification and characterization with genomic structural equation modeling. Pain. 2023;164(10):2239-52.
34. Freidin MB, Tsepilov YA, Palmer M, Karssen LC, Suri P, Aulchenko YS, et al. Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individuals. Pain. 2019;160(6):1361-73.
35. Messersmith EK, Leonardo ED, Shatz CJ, Tessier-Lavigne M, Goodman CS, Kolodkin AL. Semaphorin III can function as a selective chemorepellent to pattern sensory projections in the spinal cord. Neuron. 1995;14(5):949-59.
36. Tessier-Lavigne M, Goodman CS. The molecular biology of axon guidance. Science. 1996;274(5290):1123-33.

ARTICLE IN PRESS

ARTICLE IN PRESS

37. Gormley P, Anttila V, Winsvold BS, Palta P, Esko T, Pers TH, et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. Nat Genet. 2016;48(8):856-66.
38. Paldy E, Simonetti M, Worzfeld T, Bali KK, Vicuna L, Offermanns S, Kuner R. Semaphorin 4C Plexin-B2 signaling in peripheral sensory neurons is pronociceptive in a model of inflammatory pain. Nat Commun. 2017;8(1):176.
39. Fujii T, Yamasaki R, Iinuma K, Tsuchimoto D, Hayashi Y, Saitoh BY, et al. A Novel Autoantibody against Plexin D1 in Patients with Neuropathic Pain. Ann Neurol. 2018;84(2):208-24.
40. Hochberg MC, Carrino JA, Schnitzer TJ, Guermazi A, Walsh DA, White A, et al. Long-Term Safety and Efficacy of Subcutaneous Tanezumab Versus Nonsteroidal Antiinflammatory Drugs for Hip or Knee Osteoarthritis: A Randomized Trial. Arthritis Rheumatol. 2021;73(7):1167-77.
41. Hu Y, Shmygelska A, Tran D, Eriksson N, Tung JY, Hinds DA. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. Nat Commun. 2016;7:10448.
42. Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat Genet. 2019;51(3):394-403.
43. Ebisawa T, Uchiyama M, Kajimura N, Mishima K, Kamei Y, Katoh M, et al. Association of structural polymorphisms in the human period3 gene with delayed sleep phase syndrome. EMBO Rep. 2001;2(4):342-6.
44. Shearman LP, Jin X, Lee C, Reppert SM, Weaver DR. Targeted disruption of the mPer3 gene: subtle effects on circadian clock function. Mol Cell Biol. 2000;20(17):6269-75.
45. Daguet I, Raverot V, Bouhassira D, Gronfier C. Circadian rhythmicity of pain sensitivity in humans. Brain. 2022;145(9):3225-35.
46. Knezevic NN, Nader A, Pirvulescu I, Pynadath A, Rahavard BB, Candido KD. Circadian pain patterns in human pain conditions - A systematic review. Pain Pract. 2023;23(1):94-109.
47. Yamakawa W, Yasukochi S, Tsurudome Y, Kusunose N, Yamaguchi Y, Tsuruta A, et al. Suppression of neuropathic pain in the circadian clock-deficient Per2(m/m) mice involves up-regulation of endocannabinoid system. PNAS Nexus. 2024;3(1):pgad482.
48. Finan PH, Goodin BR, Smith MT. The association of sleep and pain: an update and a path forward. J Pain. 2013;14(12):1539-52.
49. Duffy JF, Cain SW, Chang AM, Phillips AJ, Munch MY, Gronfier C, et al. Sex difference in the near-24-hour intrinsic period of the human circadian timing system. Proc Natl Acad Sci U S A. 2011;108 Suppl 3(Suppl 3):15602-8.
50. Ong JC, Taylor HL, Park M, Burgess HJ, Fox RS, Snyder S, et al. Can Circadian Dysregulation Exacerbate Migraines? Headache. 2018;58(7):1040-51.
51. Zadina JE, Hackler L, Ge LJ, Kastin AJ. A potent and selective endogenous agonist for the mu-opiate receptor. Nature. 1997;386(6624):499-502.
52. Pokhilko A, Nash A, Cader MZ. Common transcriptional signatures of neuropathic pain. Pain. 2020;161(7):1542-54.
53. Hwang IC, Park JY, Myung SK, Ahn HY, Fukuda K, Liao Q. OPRM1 A118G gene variant and postoperative opioid requirement: a systematic review and meta-analysis. Anesthesiology. 2014;121(4):825-34.

ARTICLE IN PRESS

ARTICLE IN PRESS

54. Doyle HH, Eidson LN, Sinkiewicz DM, Murphy AZ. Sex Differences in Microglia Activity within the Periaqueductal Gray of the Rat: A Potential Mechanism Driving the Dimorphic Effects of Morphine. J Neurosci. 2017;37(12):3202-14.
55. Pisanu C, Franconi F, Gessa GL, Mameli S, Pisanu GM, Campesi I, et al. Sex differences in the response to opioids for pain relief: A systematic review and meta-analysis. Pharmacol Res. 2019;148:104447.
56. Hawkins PT, Stephens LR. PI3K signalling in inflammation. Biochim Biophys Acta. 2015;1851(6):882-97.
57. Tian Z, Shofer FS, Yao L, Sun H, Zhang H, Qin L, et al. TNFAIP8 family gene expressions in the mouse tail intervertebral disc injury model. JOR Spine. 2020;3(2):e1093.
58. Joseph EK, Levine JD. Caspase signalling in neuropathic and inflammatory pain in the rat. Eur J Neurosci. 2004;20(11):2896-902.
59. Zhang H, Li N, Li Z, Li Y, Yu Y, Zhang L. The Involvement of Caspases in Neuroinflammation and Neuronal Apoptosis in Chronic Pain and Potential Therapeutic Targets. Front Pharmacol. 2022;13:898574.
60. Zhang K, Li P, Jia Y, Liu M, Jiang J. Non-coding RNA and n6-methyladenosine modification play crucial roles in neuropathic pain. Front Mol Neurosci. 2022;15:1002018.
61. Wu L, Tang H. The role of N6-methyladenosine modification in rodent models of neuropathic pain: from the mechanism to therapeutic potential. Biomed Pharmacother. 2023;166:115398.
62. Lu W, Yang X, Zhong W, Chen G, Guo X, Ye Q, et al. METTL14-mediated m6A epitranscriptomic modification contributes to chemotherapy-induced neuropathic pain by stabilizing GluN2A expression via IGF2BP2. J Clin Invest. 2024;134(6).
63. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33(2):272-9.
64. Ni G, Moser G, Schizophrenia Working Group of the Psychiatric Genomics C, Wray NR, Lee SH. Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. Am J Hum Genet. 2018;102(6):1185-94.
65. Peterson RE, Kuchenbaecker K, Walters RK, Chen CY, Popejoy AB, Periyasamy S, et al. Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. Cell. 2019;179(3):589-603.
66. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol. 2017;186(9):1026-34.
67. Kurki MI, Karjalainen J, Palta P, Sipila TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613(7944):508-18.
68. Barakat C, Konstantinidis T. A Review of the Relationship between Socioeconomic Status Change and Health. Int J Environ Res Public Health. 2023;20(13).
69. Mayhew AJ, Meyre D. Assessing the Heritability of Complex Traits in Humans: Methodological Challenges and Opportunities. Curr Genomics. 2017;18(4):332-40.

ARTICLE IN PRESS

ARTICLE IN PRESS

70. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015;12(3):e1001779.
71. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9.
72. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, et al. Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet. 2021;53(7):1097-103.
73. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011;12:35.
74. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33(18):2938-40.
75. RStudio Team. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA; 2020.
76. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006;34(Database issue):D590-8.
77. Li S, Carss KJ, Halldorsson BV, Cortes A. Whole-genome sequencing of half-a-million UK Biobank participants. medRxiv. 2023:2023.12.06.23299426.
78. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genomic mutational constraint map using variation in 76,156 human genomes. Nature. 2024;625(7993):92-100.
79. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122.
80. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47(11):1236-41.
81. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. Genet Epidemiol. 2009;33(1):79-86.
82. Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023;622(7982):329-38.
83. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman AK, Schork A, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. Nat Metab. 2020;2(10):1135-48.
84. Pietzner M, Wheeler E, Carrasco-Zanini J, Cortes A, Koprulu M, Worheide MA, et al. Mapping the proteo-genomic convergence of human diseases. Science. 2021;374(6569):eabj1541.
85. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383.
86. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;47(9):1091-8.
87. Metzner F, Adedeji A, Wichmann ML, Zaheer Z, Schneider L, Schlachzig L, et al. Experiences of Discrimination and Everyday Racism Among Children and Adolescents With an Immigrant Background - Results of a Systematic Literature Review on the Impact of

ARTICLE IN PRESS

ARTICLE IN PRESS

Discrimination on the Developmental Outcomes of Minors Worldwide. Front Psychol. 2022;13:805941.
88. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11(4):e1004219.
89. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet. 2015;47(11):1228-35.
90. Weiner DJ, Gazal S, Robinson EB, O'Connor LJ. Partitioning gene-mediated disease heritability without eQTLs. Am J Hum Genet. 2022;109(3):405-16.
91. Herrero J, Muffato M, Beal K, Fitzgerald S, Gordon L, Pignatelli M, et al. Ensembl comparative genomics resources. Database (Oxford). 2016;2016.
92. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7.
93. Csárdi G NT, Traag V, Horvát S, Zanini F, Noom D, Müller K. igraph: Network Analysis and Visualization in R. R package version 2.0.3 ed2024.
94. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. 2024;R package version 2.56.0.
95. Sayols S. rrvgo: a Bioconductor package to reduce and visualize Gene Ontology terms. microPublication Biology. 2023.
96. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, Watt S, et al. Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell. 2016 Nov 17;167(5):1398-1414.e24.
97. de Klein N, Tsai EA, Vochteloo M, Baird D, Huang Y, Chen CY, et al. Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. Nat Genet. 2023 Mar;55(3):377-388.

ARTICLE IN PRESS

ARTICLE IN PRESS

Acknowledgements

We want to acknowledge the participants and investigators of the FinnGen and UK Biobank studies. This research has been conducted using the UK Biobank Resource under Application Number 20361. The authors would like to thank Robert Scott, Sanjay Kumar, Carolyn Buser-Doepner, Emma Laing, Lisa Mohamet, and Niall Moore for their support during this project. GSK provided funding through a direct academic collaboration agreement to both the University of Oxford and University of Cardiff for the collaboration that produced this manuscript.

Author Contributions Statement

C.E.H. performed all genetics data analysis, managed and maintained data going into the manuscript and wrote the manuscript. E.U. performed the TWAS and cell-type gene enrichment. H.A.F contributed to data analysis of SNP-based heritability estimates, genetic correlation analysis and identification of overlapping genetic loci. C.E.M and J.M.H performed gene enrichment in mouse phenotypes and cell-type and pathway enrichment and contributed to data interpretation. J.D.E supported the genetic analysis, interpretation and writing of the manuscript. L.R. provided expertise for cell-type gene enrichment. N.B performed SNP based heritability and genetic correlation analysis. L.H. supported genetic correlation analysis. J.L performed gene-based analysis. A.C. supported GWAS and finemapping. P.W. supported phenotype generation, GWAS, and provided EHR expertise. U.G . V.T., A.N., M.Z.C., and J.D. designed the phenotyping algorithm. G.Y and L.A. supported initial conceptualisation of this project and partnership between GSK, Oxford and Cardiff. C.X supported initial conceptualisation of this project and reviewed and edited previous drafts of the manuscript. C.W. led the project from Cardiff, provided expertise throughout, and contributed to data interpretation. J.D. led the phenotyping analysis, prepared all phenotyping code, and led the collaboration for GSK. M.Z.C. led the collaboration for Oxford, provided genetics and clinical expertise throughout, and jointly led the supervision of this manuscript and overall collaboration with J.D. and C.W.

Competing Interests Statement

C.E.H., C.E.M., J.M.H., J.D.E., N.B., L.H., J.L., A.C., P.W., U.G., V.T., A.N., G.Y., L.A., C.X., J.D. are employees of GSK and are shareholders of GSK stock at the time of submission. The remaining authors declare no competing interests.

ARTICLE IN PRESS

ARTICLE IN PRESS

Tables

Table 1: Summary of the overall genetic findings across each of the chronic pain phenotypes defined using analgesic use prescription data.

Grouping	Phenotype	Brief Description	Meta-analysis		UK Biobank		FinnGen		Number of independent significant signals	Number of signals not previously linked to pain	Number of previously reported signals
			# Cases	# Controls	# Cases	# Controls	# Cases	# Controls			
Pain	Chronic pain	Chronic pain patients defined as >90 days prescribed analgesics vs healthy controls (GP linkage and no chronic pain)	128985	563054	35550	156141	93435	406913	48	28	20
	Strong opioid use vs healthy	Strong opioids ever prescribed vs healthy controls (GP linkage and no chronic pain)	71837	563054	16505	156141	55223	406913	36	20	16
	Strong opioid > 90 days vs healthy	Strong opioids prescribed for more than 90 days vs healthy controls (GP linkage and no chronic pain)	31944	563054	8100	156141	23844	406913	15	8	7
	Time on analgesics Q4 vs healthy	Quartile 4 of the time on analgesics compared to participants not receiving analgesics	32474	563054	9063	156141	23411	406913	21	12	9
	Cessation Time Q4 vs healthy	Quartile 4 of the time to stop analgesics compared to participants not receiving analgesics	32412	563054	8992	156141	23420	406913	13	10	3
Pain Severity	Strong opioid use > 90 days vs never	Strong opioids ever prescribed vs patients with >90 days prescribed analgesics but never prescribed strong opioids	31915	57085	8071	18982	23844	38103	2	2	0
	Strong opioid >90 days vs <90 days	Strong opioids prescribed for more than 90 days vs Strong opioids prescribed for less than 90 days	31915	23682	8071	5291	23844	18391	0	0	0
Pain Duration	Time on analgesics Q4 vs Q1	Quartile 4 of the time on analgesics compared to Quartile 1 of the time on analgesics	32432	33074	9021	8709	23411	24365	1	1	0
	Time on analgesics Q4 vs Q1-Q3	Quartile 4 of the time on analgesics compared to all others with time on analgesics	32432	96427	9021	26403	23411	70024	2	2	0
	Cessation Time Q4 vs Q1	Quartile 4 of the time to stop analgesics compared to Quartile 1 of the time to stop analgesics	32382	32259	8962	8894	23420	23365	0	0	0
	Cessation Time Q4 vs Q1-Q3	Quartile 4 of the time to stop analgesics compared to all others time to stop analgesics	32382	96477	8962	26462	23420	70015	2	2	0

ARTICLE IN PRESS

ARTICLE IN PRESS

ARTICLE IN PRESS

ARTICLE IN PRESS

Figure Legends/Captions

Figure 1: Manhattan plot representing the genetic associations identified through a GWAS meta-analysis of chronic pain, defined using analgesic-use data, in 128,985 cases and 563,054 controls of European descent.
Chronic pain was defined using long-term analgesic-use prescription data in UK Biobank and FinnGen participants. On the x-axis, variants are plotted along the 22 autosomes included in the meta-analysis. The y-axis shows the statistical strength of association from the inverse-variance weighted fixed-effects meta-analysis as the negative $\log_{10}$ of the uncorrected P-value (*P*). The horizonal line is the genome-wide significance threshold after correction for multiple testing ($P = 5 \times 10^{-8}$). Each dot represents a single nucleotide polymorphism (SNP). Adjacent chromosomes are coloured in different shades of blue. Some well-known pain genes have been annotated by name and circle. *P* values are two-sided raw *P* values derived from a fixed-effect meta-analysis. Source data is available for download from the GWAS catalogue (https://www.ebi.ac/gwas/home, accession id: GCST90705000)

Figure 2: The overlap of genetic loci associated with the three aspects of pain defined using analgesic data - chronic pain susceptibility, pain severity, and pain duration – and in sex-stratified analyses.
Overlapping loci were determined based upon an arbitrary 1 mega-base distance between lead variants associated with any of the 11 chronic pain phenotypes. A: Upset plot representing the number of genetic loci associated with chronic pain susceptibility (red bar), pain severity (green bar), or pain duration (purple bar). Most loci are associated with chronic pain susceptibility, whilst one locus is specific to pain severity and five genetic loci are specific to pain duration. Source data: Supplementary Data 3. B: Venn diagram representing the number of genetic loci associated with chronic pain phenotypes in the joint (purple shaded circle) and sex-stratified analyses (Male analysis indicated by blue shaded circle, female analysis indicated by yellow shaded circle). There are 10 male-specific loci (representing 11 genome-wide associations) and 17 female-specific loci (representing 21 associations) associated with at least one of our 11 chronic pain phenotypes. Source Data: Supplementary Data 3 & 6. C: Upset plot representing the number of male- (indicated by green bars) or female-specific (indicated by yellow bars) associated with either chronic pain susceptibility (red bars), pain severity (green bars) or pain duration (purple bars). Most of the male- or female-specific loci, similarly to the joint analyses, are associated with pain susceptibility. Source Data: Supplementary Data 6.

Figure 3: Functional enrichment analysis highlights key cell-types, genes, and pathways implicated in chronic pain.
A) Cell-types associated with chronic pain phenotypes grouped by central and peripheral nervous systems. Associations are represented as -log10(*P* value) with the strength of association shown by the blue – red shading with the blue representing the strongest associations and red the weakest associations. Associations were identified using different methods (MAGMA, LDSC, and AMM) – full details can be seen in Supplementary Figure 3. One-sided enrichment tests were performed using MAGMA and AMM, while two-sided tests were performed using LDSC, and p-values were adjusted using the Bonferroni correction. Source Data: Supplementary Data 8. B) Gene-ontology (GO) enriched terms for oligo-precursor cells in the significantly associated modules (Modules 4 & 5). Size of dots represent the number of genes in GO terms in the oligo-precursor cell modules and the colors (red and blue) represent the degree of association with red representing the weakest associations and blue representing the strongest associations. The y-axis are the enriched GO terms. One-sided enrichment tests for gene ontology terms were

ARTICLE IN PRESS

performed with topGO, and p-values were adjusted for multiple comparisons. Source Data: Supplementary Data 9. C) Protein-protein interactions (PPI) network for genes present in both cell fate commitment and regulation of cell differentiation. Genes with a red circle are known to be associated with pain disorders from a literature search and genes with a blue circle have not previously been identified as associated with a pain disorder from a literature search. Source Data: Supplementary Data 10 D) Mouse phenotype gene sets significantly enriched in chronic pain-associated cell-types. Size of dots represents the proportion of cell-type-specific genes in each mouse phenotype gene-set and the colour represents the degree of association, with blue indicating a stronger association and red a weaker association. The top 5 mouse phenotypes are shown for each cell-type. A square around the circle indicates an additional significant association of the cell-type-specific mouse phenotype genes with one of the chronic pain GWAS. Enrichment was tested using the one-sided hypergeometric test implemented with the clusterProfiler R package and was relative to the background set of genes detected in each scRNA-seq dataset. P values were corrected using the Benjamini-Hochberg method. Source Data: Supplementary Data 11. Abbreviations: PEP, peptidergic neurons; OPC, Oligo-precursor cells; Ab RA LTMRs, rapidly adapting low-threshold mechanoreceptors; SN, Substantia nigra; TG, Trigeminal ganglia; DRG, Dorsal root ganglia; GO, Gene Ontology; A, Combined analysis of males and females.

Figure 4: Gene-set overlap and pathway analyses

A) Upset plot representing the overlap between genes identified through our complementary approaches, including GWAS, rare-variant gene-based burden testing, TWAS as well as integrating known pain genes identified by Li et al. 2023 [7]. Source Data: Supplementary Data 16. B) The top four most significant GO terms enriched in our gene-set of all pain genes identified through GWAS, gene-based burden testing, and TWAS alongside inclusion of pain genes. Genes identified through TWAS are coloured in blue and directionality inferred for these genes from TWAS is represented by a '+' or '-' in brackets. '+' indicates that increased gene-expression is associated with increased disease risk whilst '-' indicates that increased gene expression is associated with decreased disease risk. Known pain genes are coloured Red. Source Data: Supplementary Data 17. C) Summary of the direction of effects inferred from TWAS for some genes identified through TWAS. Abbreviations: TWAS, transcriptome-wide association study; CTS, Cell-type-specific; TN, tibial nerve; ODC, Oligodendrocytes; SO vs H, Strong opioid vs healthy; CP vs H, chronic pain vs healthy; OPC, Oligo-precursor cells. Source Data: Supplementary Data 15.

Figure 5: Heat map of the shared genetic effects between chronic pain phenotypes and other pain-related, immune-related and psychiatric conditions.

Heat map shows only those results passing a Bonferroni corrected $P < 4.8 \times 10^{-5}$ (0.05 / 1037 independent tests) and with a genetic correlation score > |0.5|. The y-axis shows the chronic pain phenotypes defined using the prescription data clustered by pain susceptibility, pain duration and pain severity. The x-axis shows the phenotypes for which we were assessing genetic correlation clustered by pain, and psychiatric. The darker the orange shading the stronger the genetic correlations. There were no significant genetic correlations between chronic pain phenotypes and immune or anthropometric traits. Genetic correlation was performed using linear regression

ARTICLE IN PRESS

implemented in LDSC. Statistical tests were two-sided. Source Data: Supplementary Data 20. Abbreviations: DpP, depression with chronic pain; -DpP, no depression or chronic pain; PTSD, Post-traumatic stress disorder; 90d, 90 days; OA, osteoarthritis; meta, meta-analysis; Q4, quartile 4; Q1, quartile 1.

Figure 6: Overview of the analyses performed to identify genes and pathways implicated in chronic pain.

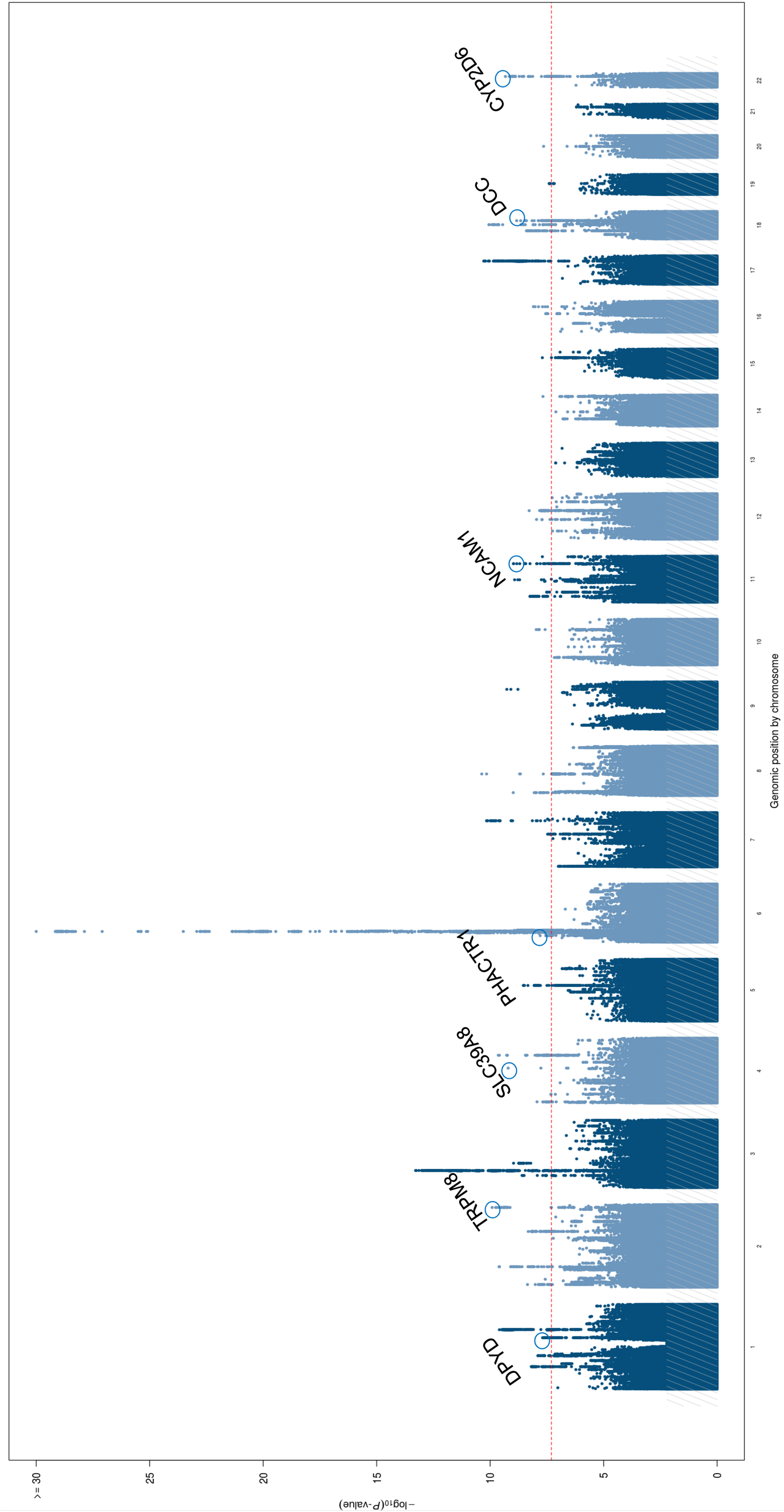
Briefly, we used analgesic use prescription data in electronic health records to define 11 chronic pain phenotypes in UK Biobank and FinnGen. Using these 11 chronic pain phenotypes, we performed GWAS of each phenotype in both cohorts, followed by meta-analyses to identify common variants associated with chronic pain. The most likely causal genes in each GWAS associated locus were prioritised based on a series of complementary computational approaches, including nearest gene, e/pQTL colocalisations and the presence of high or moderate impact functional variants in the credible set. We investigated the genetic overlap with other pain-related conditions as well as other conditions comorbid with pain, such as psychiatric and immune-related. The genetic associations were then integrated with orthogonal types of data, including cell-type gene expression data, pathway enrichment, and mouse phenotype enrichment, to provide additional insights into underlying biology and pathophysiology of pain. Source Data: Supplementary Data 1-2, 4, 7, 21, 23, 24. Abbreviations: eQTL = expression quantitative trait loci, pQTL = protein quantitative trait loci, N = Sample size, Rx = prescription.

ARTICLE IN PRESS

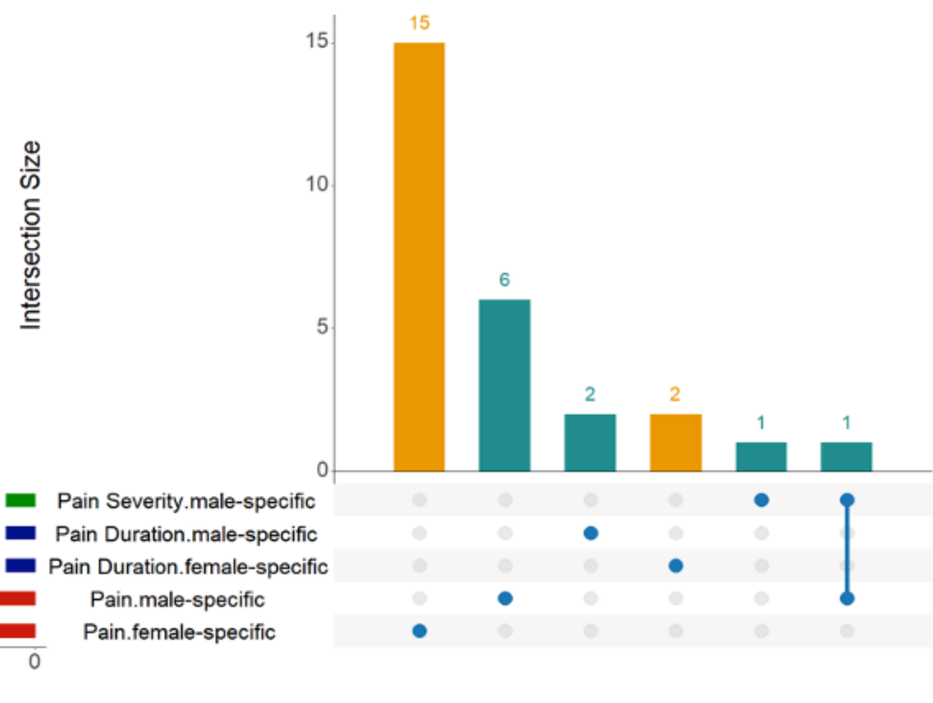
Editorial Summary

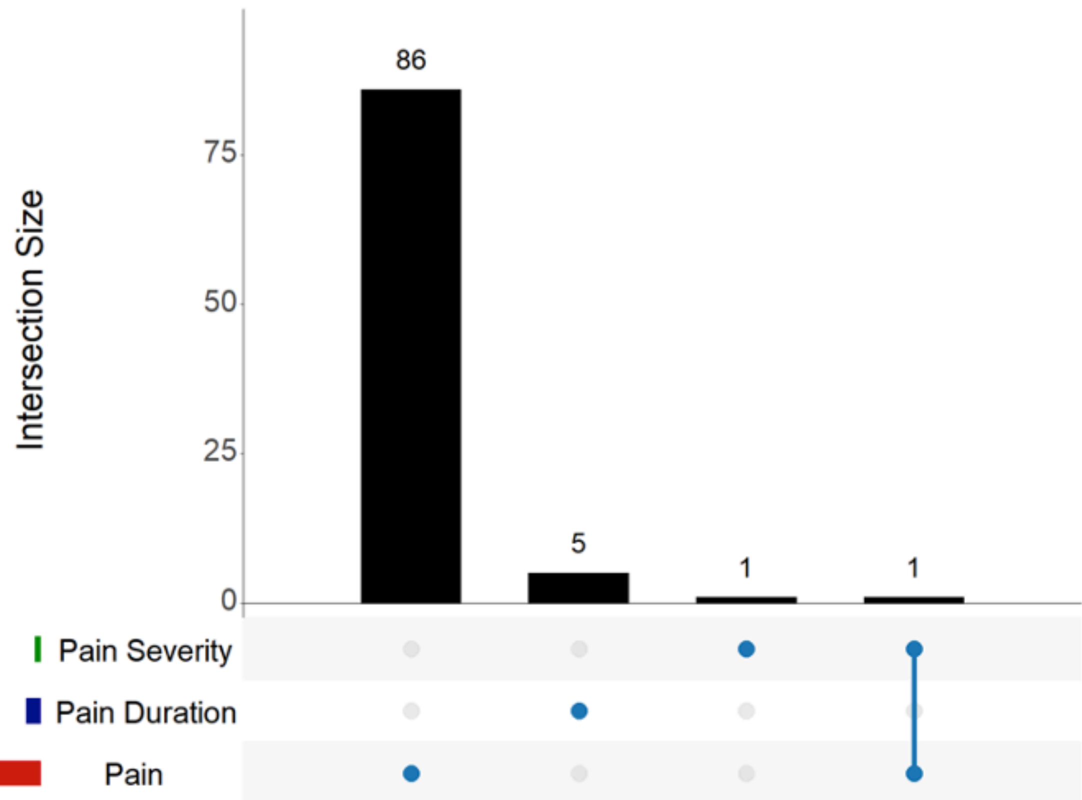
Analyses of chronic pain have traditionally focused on people with pain-related disorders, excluding those without a specific condition. To avoid this, Harlow et al. perform a GWAS meta-analysis of chronic pain defined based on the use of analgesics.

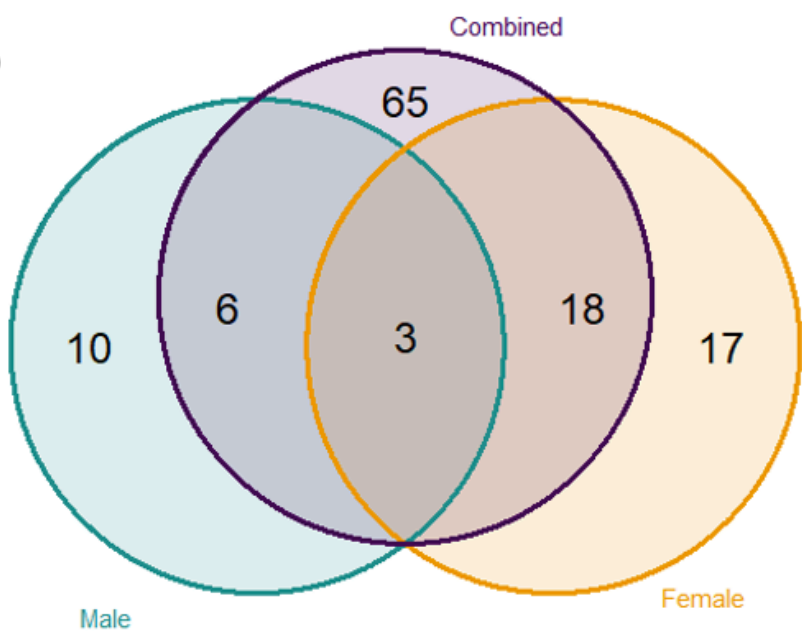
Peer Review Information: *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

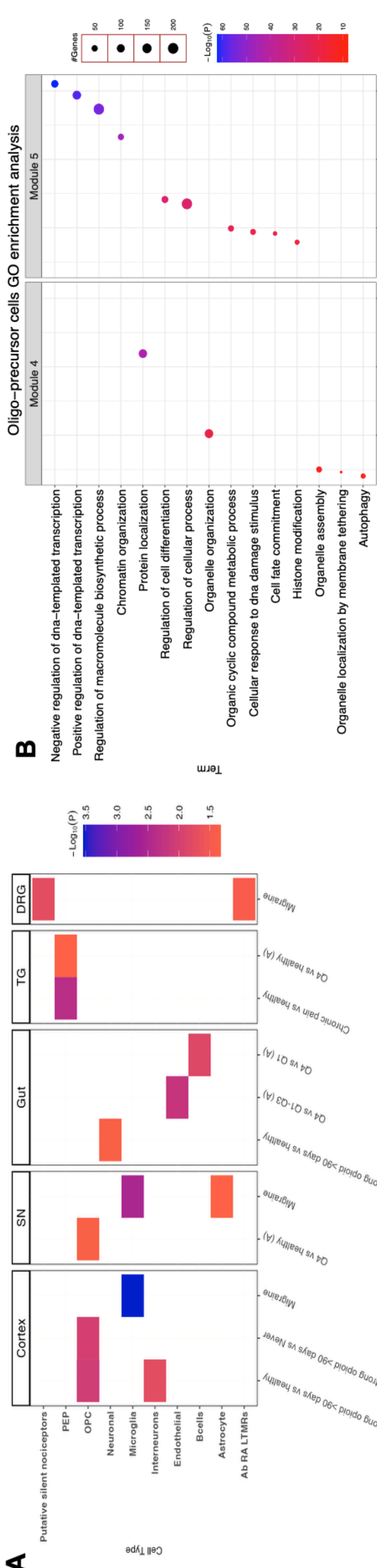


DPYD
TRPM8
SLC39A8
PHACTR1
NCAM1
DCC
CYP2D6
Genomic position by chromosome
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
− log10(P-value)
0 5 10 15 20 25 >= 30


A
86
5
1
1
Intersection Size
75
50
25
0
Pain Severity
Pain Duration
Pain
75 50 25 0
Set Size

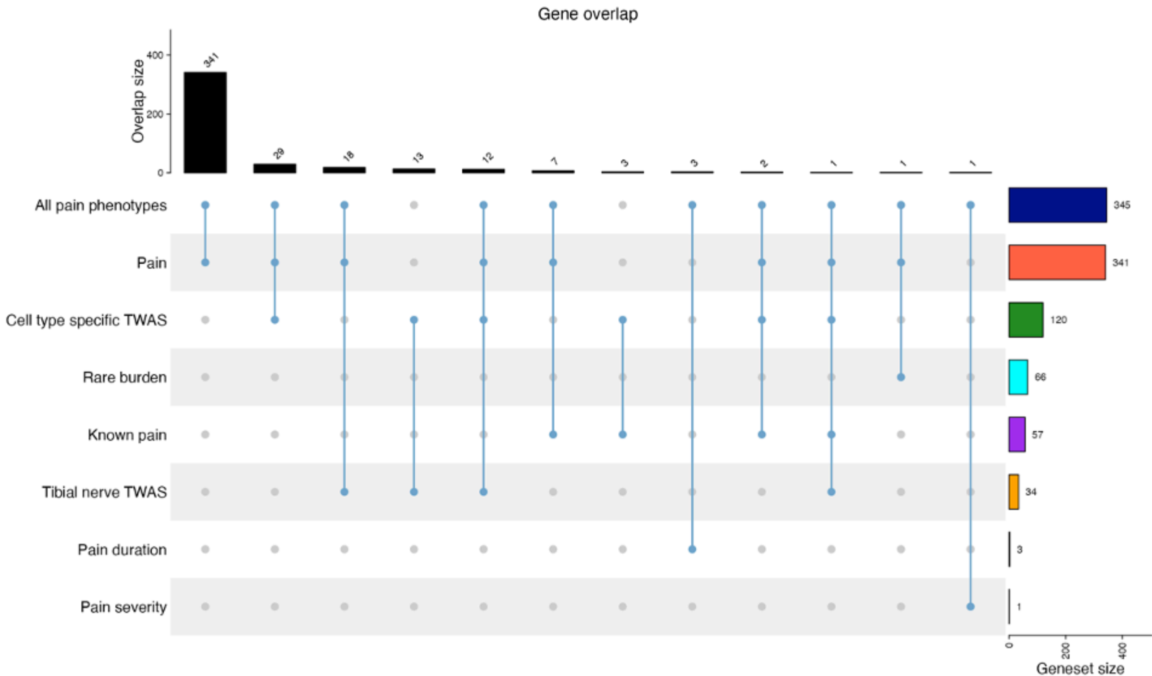

B
Combined
Male
Female
10
6
65
3
18
17


C
Intersection Size
15
10
5
0
15
6
2
2
1
1
Pain Severity.male-specific
Pain Duration.male-specific
Pain Duration.female-specific
Pain.male-specific
Pain.female-specific
15 10 5 0
Set Size


A
Cortex
SN
Gut
TG
DRG
Cell Type
Putative silent nociceptors
PEP
OPC
Neuronal
Microglia
Interneurons
Endothelial
Bcells
Astrocyte
Ab RA LTMRs
Strong opioid >90 days vs healthy
Strong opioid >90 days vs Never
Migraine
Q4 vs healthy (A)
Q4 vs Q1-Q3 (A)
Q4 vs Q1 (A)
Chronic pain vs healthy
Trait
-Log10(P)
C
OLIG2
NKX2-2
ASCL1
NOTCH2
HES1
MITF
RBPJ
HDAC1
TOX
SOX6
SOX5
SOX2
SMAD2
CDC73
TCF7L2
SOX13

B
Oligo−precursor cells GO enrichment analysis
Module 4
Module 5
Negative regulation of dna−templated transcription
Positive regulation of dna−templated transcription
Regulation of macromolecule biosynthetic process
Chromatin organization
Protein localization
Regulation of cell differentiation
Regulation of cellular process
Organelle organization
Organic cyclic compound metabolic process
Cellular response to dna damage stimulus
Cell fate commitment
Histone modification
Organelle assembly
Organelle localization by membrane tethering
Autophagy
Term
−Log10(P)
#Genes
50
100
150
200
D
Cortex
SN
Gut
TG
DRG
Neonatal lethality, complete penetrance
Failure to thrive
B lymphocytopenia
Decreased CD4-positive, alpha beta T cell number
Leukocytosis
Abnormal T cell differentiation
Abnormal B cell morphology
Decreased body size
Abnormal vitelline vasculature morphology
Abnormal macrophage physiology
Abnormal bleeding
Peripheral demyelination
Splenomegaly
Abnormality of B cell physiology
Abnormal excitatory postsynaptic currents
Abnormal nervous system electrophysiology
Ataxia
Gait disturbance
Seizure
Abnormal spatial working memory
Small cerebellum
Hyperactivity
Limb grasping
Abnormal axon extension
Respiratory distress
Increased neutrophil cell number
Abnormal humoral immune response
Decreased circulating total IgM
Increased circulating IgM level
Neonatal lethality, incomplete penetrance
Decreased angiogenesis
Postnatal lethality, complete penetrance
Abnormal CNS synaptic transmission
Abnormal axon morphology
Abnormal vitelline vascular remodeling
Abnormal vascular endothelial cell morphology
Axonal degeneration
Abnormal myelin sheath morphology
Abnormal axonal transport
Interneurons (1117)
Microglia (1142)
OPC (1136)
Astrocyte (1212)
Microglia (1135)
OPC (1152)
Bcells (1114)
Endothelial (1354)
Neuronal (1269)
PEP (1238)
Ab RA LTMRs (844)
Putative silent nociceptors (840)
−Log10(P)
GeneRatio
0.05
0.10

A



B

All pain phenotype

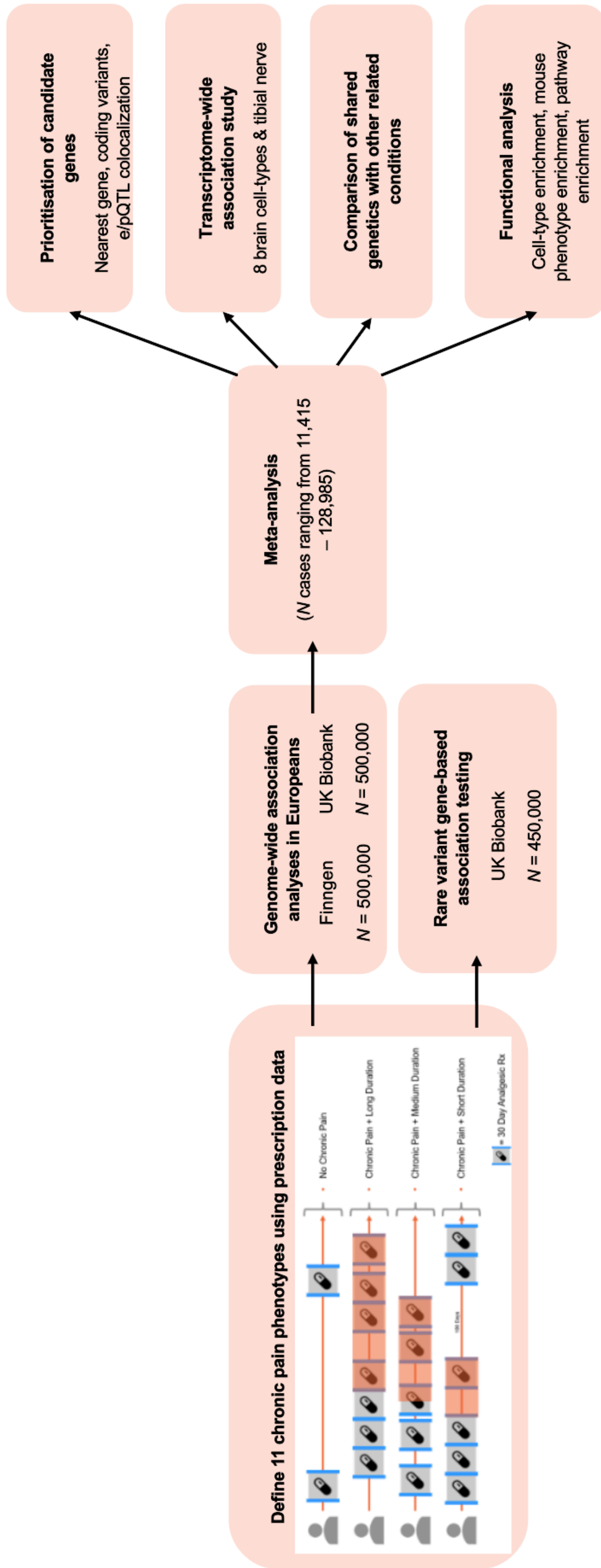
GO Terms	Count	Category	Genes: *known pain genes (Red) & TWAS genes (Blue)*	#Loci	#Shared loci	#Distinct loci
Negative regulation of neuron projection development	7	BP	(+/-) *CERS2* , *FAT3*, *DCC*, *SEMA3F*, *SEMA3B*, *AMIGO3*, *PRAG1*	11	8	3
Endolysosome	18	CC	(-) *CTSK*, *CTSS*, *CRYAB*, *CTSW*, *LRP1*, *PIP4K2C*, *RMC1*, (-) *NPC1*, *NAGA*, *HYAL2*, *HYAL1*, *NPRL2*, (+/-) *USP4*, *HYAL3*, *SLC7A14*, (-) *MST1R*, *SLC39A8*, *HTT*	16	9	7
Postsynaptic recycling endosome	20	CC	(-) *CTSK*, *RABGAP1L*, *ZFYVE9*, *CTSS*, *SNX32*, *LRP1*, *DDIT3*, *RMC1*, (-) *NPC1*, *PHETA2*, (+) *RHOA*, *HYAL3*, *MON1A*, *HTT*, *OPRM1*, *BLTP3A*, *CNTNAP2*, *DYNC1I1*, *ZDHHC2*, *ASTN2*	24	14	10
Plasma membrane region	7	CC	*CSNK1G2*, *CLIP4*, *SEPTIN3*, (+) *RHOA*, (+) *BSN*, *PPP1R9A*, *ASTN2*	13	10	3

C

Gene	TWAS	Direction	Phenotype	Cell
CERS2	CTS TN	+ -	CP vs H	Endothelial -
CTSK	CTS	-	CP vs H	Inhibitory
MST1R	TN CTS	- -	CP vs H & SO v H SO vs H	- Inhibitory, Excitatory
NPC1	TN	-	CP vs H	-
USP4	CTS	+ -	CP vs H SO vs H	Microglia, Inhibitory Pericytes, OPC
RHOA	CTS	+	SO vs H	Excitatory
BSN	CTS	+	CP vs H, SO v H	Inhibitory
ARNT	CTS	+	CP vs H	Microglia
AMT	TN	-	SO vs H	-
DALRD3	TN CTS	- -	SO vs H	- ODC, Excitatory

Genetic Correlation
0.6
0.7
0.8

Pain
Pain Duration
Pain Severity
Chronic Pain
Analgesic Time
Cessation Time
Strong Opioid
Analgesic Time
Cessation Time
Strong Opioid
Psychiatric
PTSD
Depression
Anxiety
Pain
Oesteoarthritis
Migraine
Fibromyalgia
DpP vs -DpP
Chronic Fatigue
Chronic Fatigue
PTSD (Stein et al. 2021) - Total PCL
PTSD (Stein et al. 2021) - Avoidance symptoms
PTSD (Stein et al. 2021)
PTSD (Nievergelt et al. 2019)
PTSD (Gelernter et al. 2019)
Depression meta (UKB & Levey et al. 2021)
Depression (Levey et al. 2021)
Anxiety (Levey et al. 2020)
OA meta (Finngen & Boer et al. 2021)
Knee OA meta (Finngen & Boer et al. 2021)
Knee OA (Boer et al. 2021)
Hip OA (Boer et al. 2021)
Migraine meta (UKB & Finngen)
Fibromyalgia meta (Finngen & Estonia BB)
Depression with Pain vs No Pain or Depression (UKB)
Chronic Fatigue (UKB)
Chronic Pain Vs Healthy
Q4 vs Healthy
Q4 vs Healthy
Strong Opioid vs Healthy
90d vs Healthy
Q4Q1
Q4 v all
Q4Q1
90d vs Never
90d vs <90


Define 11 chronic pain phenotypes using prescription data
No Chronic Pain
Chronic Pain + Long Duration
Chronic Pain + Medium Duration
Chronic Pain + Short Duration
= 30 Day Analgesic Rx
Genome-wide association analyses in Europeans
Finngen
N = 500,000
UK Biobank
N = 500,000
Rare variant gene-based association testing
UK Biobank
N = 450,000
Meta-analysis
(N cases ranging from 11,415 – 128,985)
Prioritisation of candidate genes
Nearest gene, coding variants, e/pQTL colocalization
Transcriptome-wide association study
8 brain cell-types & tibial nerve
Comparison of shared genetics with other related conditions
Functional analysis
Cell-type enrichment, mouse phenotype enrichment, pathway enrichment