Genome-wide analysis identifies a distinct pathophysiology for chronic overlapping pain conditions via impaired axonogenesis in the brain

Samar Khoury[1, †], Marc Parisien[1, †], Scott J. Thompson[1,2], Etienne Vachon-Presseau[1], Mathieu Roy[1,3], Amy E. Martinsen[4,5,6], Bendik S. Winsvold[4,5,7], the HUNT All-In PAIN group[8], Ingunn Mundal[9], John-Anker Zwart[4,5,6], Artur Kania[10], Jeffrey S. Mogil[1,3], Luda Diatchenko[1, *]

1 Alan Edwards Centre for Research on Pain; Faculty of Dentistry; Department of Anesthesia, Faculty of Medicine, McGill University, Montréal, QC, Canada
† Equally contributing authors
2 Department of Anesthesiology, University of Minnesota, Minneapolis, MN USA
3 Department of Psychology, McGill University, Montréal, QC, Canada
4 K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway
5 Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway
6 Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
7 Department of Neurology, Oslo University Hospital, Oslo, Norway
8 HUNT All-In Pain Consortium
9 Department of Health Science, Molde University College, Molde, Norway
10 Institut de recherches cliniques de Montréal (IRCM); Department of Cell Biology and Anatomy, and Experimental Medicine, McGill University, Montréal, QC, Canada

* Corresponding author:
Professor Luda Diatchenko, MD, PhD
Faculty of Dentistry
Department of Anesthesia, Faculty of Medicine
McGill University
Genome Building, Room 2201
740 Dr. Penfield Avenue
Montreal, Quebec, Canada H3A 0G1
T: 514 398-2878 F: 514 398-8900

This is an accepted, peer reviewed postprint-version of the following journal article that has been published in the journal Brain: Samar Khoury, Marc Parisien, Scott J Thompson, Etienne Vachon-Presseau, Mathieu Roy, Amy E Martinsen, Bendik S Winsvold, HUNT All-In Pain, Ingunn P Mundal, John-Anker Zwart, Artur Kania, Jeffrey S Mogil, Luda Diatchenko, Genome-wide analysis identifies impaired axonogenesis in chronic overlapping pain conditions, Brain, Volume 145, Issue 3, March 2022, Pages 1111–1123. DOI: https://doi.org/10.1093/brain/awab359

ABSTRACT

Chronic pain is often present at more than one anatomical location, leading to chronic overlapping pain conditions (COPC). Whether COPC represents a distinct pathophysiology from the occurrence of pain at only one site is unknown. Using genome-wide approaches, we compared genetic determinants of chronic single-site vs. multi-site pain in the UK Biobank. We found that different genetic signals underlie chronic single-site and multi-site pain with much stronger genetic contributions for the latter. Among 23 loci associated with multi-site pain, 9 loci replicated in the HUNT cohort, with the DCC netrin-1 receptor (*DCC*) as the top gene. Functional genomics identified axonogenesis in brain tissues as the major contributing pathway to chronic multi-site pain. Finally, multimodal structural brain imaging analysis showed that *DCC* is most strongly expressed in subcortical limbic regions and is associated with alterations in the uncinate fasciculus microstructure, suggesting that DCC-dependent axonogenesis may contribute to COPC via cortico-limbic circuits.

Chronic pain is a common and complex disease with a prevalence of 10–50% worldwide and is associated with substantial costs to affected individuals and society at large[1-3]. The clinical assessment of most chronic pain conditions relies on self-report of symptoms associated with a specific anatomical location. However, at least one-third of chronic pain patients diagnosed with one pain condition often simultaneously exhibit symptoms of another[4,5]. Epidemiological studies have examined the overlap between different bodily distribution of pain and suggested that they may share a common underlying etiology[5]. In these pain conditions, recently referred to as nociplastic, altered network architecture of functional brain connectivity seems to contribute to central sensitization and co□occurring symptoms include fatigue, mood and cognitive problems, sleep disturbances, and multisensory hypersensitivity[6]. The most common set of pain disorders that tend to overlap includes temporomandibular disorders, fibromyalgia, irritable bowel syndrome, vulvodynia, myalgic encephalomyelitis/chronic fatigue syndrome, headaches, and chronic lower back pain. This manifestation of multiple chronic pain conditions that frequently occur together and are associated with similar risk factors are referred to as chronic overlapping pain conditions (COPC), and are now recognized by the National Institute for Health (NIH) as a set of disorders that co-occur[7]. Although the pathophysiological processes that underlie most of these conditions are still poorly understood, COPC have been proposed to have common genetic, neurological, and psychological vulnerabilities.

Twin studies have indicated that chronic pain conditions show a heritability between 16–50%[8]. Shared heritability between pelvic pain and facial pain, and between widespread pain and abdominal pain have been reported[9,10]. Candidate gene studies have suggested that the same genetic variants are associated with multiple pain conditions, which implicated a possible shared genetic basis[11]. There remains a paucity of genetic findings based on genome-wide association

studies (GWAS) in large cohorts that have systematically assessed multiple chronic pain conditions. To date, most genetic association studies of pain have featured small samples of a single pain condition, with a few exceptions for back pain and multi-site pain[12,13]. It is still unknown whether the reports of COPC versus one specific chronic pain condition feature distinct pathophysiologies or are simply a manifestation of one another.

In this study, we employed genome-wide and brain structure analysis to understand the pathophysiology of COPC. Our first objective was to understand the genetic basis of chronic pain manifestation at one body site versus multiple body sites as a proxy for COPC. Our second objective was to uncover the molecular pathophysiology underlying COPC. Our final objective was to investigate whether central nervous system (CNS) mechanisms are genetically related to COPC. Our goal was to uncover the shared genetic heritability between chronic pain conditions and to search for potential underlying biological pathways for COPC.

RESULTS

Prevalence of chronic pain sites

In the UK Biobank, 294,627 participants (60%) reported pain that interfered with their usual activities in the past month. Participants were given the choice among eight pain sites, with the possibility to report more than one site (Figure 1A): head, facial, neck/shoulder, back, stomach/abdominal, hip, knee, and "all over the body". The highest prevalence reported was for back (26%) and neck/shoulder (23%) pains. These participants were then asked if their pain lasted for more than three months. Participants who answered "yes" for pain that lasted for more than three months were classified as having chronic pain. Participants reported chronic pain for at least one site at 72%. The highest prevalence of chronic pain was reported for back (18%), knee (17%), and neck (16%) pains. Headache (9%), hip (9%), and abdominal (5%) pains showed less than 10% prevalence. Pain all over the body (1%) and facial pain (1%) displayed the lowest prevalence. Participants that reported pain in the last month and for more than three months at the same site, were defined as having pain chronification. Pain all over the body, knee, and hip pains showed the highest rates of chronification (81%, 78%, and 77%, respectively; Supplementary Table 1).

Next, we created two distinct groups to represent participants who reported only one chronic pain site and those who reported pain at two or more pain sites, which include participants with pain all over the body. We defined participants who reported more than one pain site for more than three months as participants with multi-site pain as a proxy for COPC. One third (34.1%) of participants with chronic pain reported multi-site pain and 38% reported single-site pain. Around 28% of participants did not report any chronic pain site (Supplementary Figure 1). In participants with multi-site pain, the highest odds ratio (OR) for pain at two sites

was for facial pain and headache (OR [95%CI] =10.7 [10.1-11.5]), followed by back and hip pain (OR [95%CI] =5.9 [5.8-6.1]) (Figure 1B, Supplementary Table 2). Pain all over the body was excluded from this analysis because participants who indicated pain all over the body did not have the option to report any other pain site. Participants who reported multi-site pain were more likely to be older, female, have higher body mass index, and have lower socioeconomic status. They were also more likely to report more cancer and non-cancer illnesses and to consume more paracetamol and ibuprofen, but not aspirin. In terms of mental health status, participants with multi-site pain reported higher neuroticism scores, and a higher number of and more severe depressive episodes (Table 1).

Genetic correlation of chronic pain sites

Most chronic pain sites were found to be genetically correlated (Figure 1B, Supplementary Table 3). The largest genetic correlation was observed between facial and abdominal pain (r_g = 1.04, P=1.8x10^{-10}), followed by pain all over the body and abdominal pain (r_g= 0.99, P=8.2x10^{-8}). Headaches presented the smallest genetic correlations with any other chronic pain sites (r_g between 0.37 and 0.54). In a latent causal variable analysis to infer causality, we detected evidence for genetically causal effect of facial pain on hip pain. We also detected a genetic causal effect of headache on back, knee and neck/shoulder pains (Supplementary Table 4).

Pain site pairs that are physically close displayed stronger correlations (Figure 1B). Close physical proximity between two pain sites yields an increased chance of their being reported together (% variance explained: r^2=54%, P=1.4x10^{-4}) (Figure 1C). Also, increased genetic correlation is observed with close physical proximity (r^2=15%, P=4.9x10^{-2}) (Figure 1D). Genetic and epidemiological variables (pain sites) were also observed to be correlated (r^2=16%, P=4.7x10^{-2}) (Figure 1E).

Heritability of chronic pain sites

For each chronic pain site, we calculated the heritability derived from genome-wide association (h^2_g), defined as the proportion of phenotypic variance explained by common single nucleotide polymorphisms (SNPs) under an additive model of inheritance. Between 1–10% of the heritability can be explained for each pain site (Figure 1F, Supplementary Table 5). The highest heritability was identified for back pain (h^2_g=10.0%, P=7x10^{-106}) while the lowest was for facial pain (h^2_g=1.4%, P=1x10^{-5}).

Genome-wide associations of chronic overlapping pain conditions

Next, we performed a comparative GWAS analysis for the report of chronic single-site pain with the report of chronic multi-site pain. In a total sample of 340,547 participants we conducted a GWAS contrasting the report of one pain site (n=93,964) with a randomly selected half of participants who reported no pain at any site (n=81,805). We also conducted a GWAS contrasting the report of multi-site (n=82,812) with non-overlapping controls as the rest of the randomly selected participants who reported no pain at any site (n=81,966).

We then computed the percentage of variance explained by genetic and by environmental factors for the report of single-site versus multi-site pain. We found a substantial contribution of environmental factors for both the report of single-site (93.2%; standard error of the mean (s.e.m) 0.4%) and multi-site (80.9%; s.e.m 0.4%) pain. However, we found a significant difference (P<2.2x10^{-16}) for genetic factors between the report of single- site pain (6.9%; s.e.m.0.4%) and the report of multi-site pain (19.1%; s.e.m 0.4), with a much greater genetic contribution in chronic multi-site pain (Figure 1F). Importantly, the heritability for multi-site pain was twice higher than heritability for any individual pain site.

In the case-control association study, where cases were defined as participants reporting chronic single-site pain (n=93,964), and controls being participants not reporting any pain site (n=81,805), there were no individual loci that passed the threshold of genome-wide significance (Figure 2A, Supplementary Table 6). The genomic inflation factor lambda was 1.07, but the LD score regression intercept value was 1.015, suggesting a polygenic signal rather than inflation from unaccounted population stratification (Supplementary Figure 2A). A gene-level association analysis in MAGMA testing for 18,220 genes showed that 11 genes passed multiple testing (Bonferroni threshold $P<2.7\text{x}10^{-6}$) (Supplementary Table 7).

In the case-control genome-wide association study, where cases were defined as participants reporting chronic multi-site pain (n=82,812), and controls being participants not reporting any pain site (n=81,966), there were 896 SNPs spanning 23 loci that passed the genome-wide threshold (Figure 2B, Supplementary Figure 3, Supplementary Table 8). The genomic inflation factor lambda was 1.20, but the LD score regression intercept value was 1.017, suggesting again, a contribution of LD structure of associated loci rather than inflation from unaccounted-for population stratification (Supplementary Figure 2B). A gene-level analysis showed that 97 genes passed multiple testing ($P=2.7\text{x}10^{-6}$). The two top associations were with genes involved in neuronal connectivity in model animals: *DCC*[14], encoding the DCC receptor for netrin1 ($P=7.4\text{x}10^{-19}$), and *SDK1*[15], encoding the sidekick cell adhesion molecule 1 ($P=5.4\text{x}10^{-18}$) (Supplementary Table 9). Since both GWASs were equally powered, the differences observed at both the SNP and the gene-level analyses might partially account for the differences in heritability estimates, establishing distinct genetic backgrounds.

Genome-wide meta-analysis

In order to identify loci that were specific to individual pain states (i.e., single-site and multi-site pain) and pleiotropic loci that contribute to both states, we performed two meta-analyses using GWAMA[16]. The first meta-analysis aimed to identify loci that are distinct for each of the GWASs (Figure 2C). Out of the 18,066 genes tested, 41 genes passed the threshold for multiple testing (Supplementary Table 10). The top two genes shown in the meta-analysis are *DCC* and *SDK1*, which are also the top two genes in chronic multi-site pain. The second meta-analysis aimed to identify loci that are pleiotropic between the report of single-site pain and multi-site pain by running a classical fixed-effect meta-analysis between the two GWASs (Figure 2D). There are 36 genes that passed the threshold for multiple testing, with the top two genes being *BBX* and *PABPC4* (Supplementary Table 11). Overall, we found that there are both distinct and common genetic loci underlying chronic single-site pain and chronic multi-site pain.

Tissue-expression based functional analyses

Next, we performed partitioned heritability analyses by means of a stratified LD score regression[17,18] to examine whether the observed heritability was enriched in any tissue, regulatory region or functional category[19]. Analyses in a wide range of tissues and cell types[20] were done for both the report of single-site pain and multi-site pain. Partitioned heritability analysis for single-site pain did not show any enrichment in any of the tested tissues at a 10% false discovery rate (FDR) (Figure 3A – Top panel, Supplementary Table 12). The analysis of a wide range of tissues and cell types for chronic multi-site pain yielded significant results exclusively in the CNS, but not in other tissue types like adipose, blood or immune, and connective or musculoskeletal, nor in the peripheral nervous system (Figure 3A – Bottom panel, Supplementary Table 13). We found an exclusive significant enrichment in most brain tissues (Figure 3B). Finally, in order to quantify whether the enrichment was exclusive to multi-site

pain, we correlated the heritability estimates in brain-specific tissues. We found no evidence for tissue-based congruency between the two heritability estimates, which suggests distinct tissue heritability (Figure 3C). Tissue-expression based analysis concluded that heritability for chronic multi-site pain, and not chronic single-site pain, is exclusively enriched in the CNS.

Pathway-based functional analyses

We next performed pathway-based enrichment analyses from SNPs in gene sets using Gene Ontology's (GO) biological processes for both chronic single-site pain and multi-site pain. For the report of chronic single-site pain, there was no enrichment in any pathway at FDR 10% in GO biological process (Supplementary Table 14). For the report of chronic multi-site pain, a total of 60 pathways were significant at the FDR 10% level in GO biological process, with most pathways involved in neural development, that include *DCC* and *SDK1* as leading-edge genes (Supplementary Table 15). We then used reviGO[21] to reduce redundancy and extricate meaningful information regarding biological processes. The top reviGO class of pathway identified regulation of nervous system development that encompasses pathways involving neurogenesis, axonal development and post-synaptic specialization. Taken altogether, our pathway analysis results were in line with tissue-expression based functional analysis suggesting that pathways acting in the CNS in general and associated with neural development in particular, contribute to the pathophysiology of chronic multi-site pain. Moreover, pathway analysis further supported a strong genetic basis for chronic multi-site pain but not for chronic single-site pain.

Replication of genome-wide loci in an independent cohort

Next, we attempted to replicate the genome-wide significant SNPs in the independent HUNT cohort. Due to the absence of genome-wide significant SNPs in the chronic single-site pain GWAS, we only replicated the chronic multi-site pain variants. We attempted the replication of

the lead SNP in each of the loci and for SNPs that are in medium (r^2>=0.5) and high LD (r^2>=0.8) with it in the HUNT cohort. Out of the 23 loci, nine loci reached nominal significance at P≤0.05, of which four reached statistical significance at P≤0.002 (corrected for 23 tests) (Supplementary Table 16). The following four loci passed the threshold for multiple testing. Locus 4, with lead SNP rs11709734, located on chromosome 3 in the inositol hexakisphosphatase kinase 1 (*IP6K1*) gene. Locus 8, with lead SNP rs34595097, located on chromosome 4 in the mastermind like transcriptional coactivator 3 (*MAML3*) gene. Locus 11, with lead SNP rs12672683, located on chromosome 7 in the forkhead box P2 (*FOXP2*) gene. Finally, locus 20, with lead SNP rs8099145, located on chromosome 18 in the *DCC* gene, showed the most robust replication (P=2.0x10^{-4}) (Table 2a).

Next, we attempted to replicate the 97 genes associated with chronic multi-site pain in the UK Biobank within the HUNT cohort. The threshold for replication was corrected for 97 tests and set at P=5.6x10^{-4}. Out of the 97 genes, 11 genes successfully replicated. The most striking association is with the *DCC* gene with a p-value of 2.6x10^{-8}, reaching genome-wide statistical significance (Supplementary Table 16, Table 2).

Finally, at the pathway level, we attempted to replicate the pathways that passed FDR 10% in the UK Biobank. The axonogenesis pathway (GO:0007409) showed the lowest *P*-value in the HUNT cohort. This pathway represents mechanisms involved in *do novo* generation of axons, including the terminal branched region. This morphogenesis also includes the shape and form of the developing axon. The second pathway was axon development (GO:0061564), which covers processes that involve axon regeneration or regrowth after loss or damage (Supplementary Table 16, Table 2).

In summary, the replication of our results in HUNT cohort provided further evidence that axonogenesis through the netrin receptor DCC is important in the pathophysiology of chronic multi-site pain.

Functional validation for the role of *DCC* in the human brain

Chronic multi-site pain-related heritability seems to be expressed in brain tissues with a significant role for the axonogenesis pathway through the *DCC* gene. We therefore attempted to localize where *DCC* is most strongly expressed using a fine-grained representation of genomic information across the human brain and identify the location of axonal structures using diffusion weighted imaging.

First, normalized *DCC* expression information was obtained from approximately 500 brain samples (per hemisphere) of six deceased human donors from the Allen Human Brain Atlas[22]. A heat map representing the normalized *DCC* expression across the donors was generated using the neurosynth platform. We observed that *DCC* is specifically expressed in subcortical limbic regions, such as the hippocampus, and basal ganglia (Figure 4A-B), the corticolimbic system involved in motivation and affect regulation as well as the amplification and the chronification of pain.

Given our findings on the role of *DCC*-driven axonogenesis in chronic multi-site pain and *DCC* expression in corticolimbic circuits, we next examined the associations between the microstructure of the uncinate fasciculus (UF) which connects the prefrontal cortex to limbic structures of the temporal lobe such as the amygdala and the hippocampus (Figure 4C). The UF is also the main cortico-limbic tract available as an imaging derived phenotype (IDP) in the UK Biobank. Analyses of the UF were performed on 5378 participants that consistently reported no

pain (n=3,985), single-site pain (n=593), or multi-site pain (n=800) on both the initial visit and the brain imaging visit (about 10 years apart). Orientation dispersion (OD), a spatial organization metric that characterizes angular variation of neurites (dendrites and axons), was extracted as a metric with potential relevance to axon guidance for the left and the right uncinate fasciculus and was compared between the groups. Our analysis revealed that participants with multi-site pain showed significantly higher OD in UF compared to single-site pain and healthy controls (Figure 4D), indicating that UF white matter tracts in patients with COPC are less structured.

In order to assess whether genetic variants in *DCC* and axonogenesis pathway contribute to the OD of the UF, we generated a polygenic risk score (PRS) using summary statistics of single-site pain, multi-site pain, the axonogenesis pathway, and the *DCC* gene using the best PRS, i.e. which explains the highest variance. Each of the four scores were used as dependent variables in a regression model with left and right OD of the UF as independent variable (Supplementary Table 17). The score generated using *DCC* showed the highest significance for both brain sides OD of the UF. The PRS derived from the single-site GWAS at a *P*-value threshold of $5x10^{-8}$ explained 0.034-0.044% of the variability (P=$1.0x10^{-5}$; P=$5.5x10^{-4}$) for the left and right UF respectively. PRS derived from the multi-site pain GWAS at a *P*-value threshold of $4x10^{-2}$ explained 0.035% and 0.029% of the variability (P=$4.8x10^{-4}$; P=$1.4x10^{-3}$) for the left and right UF respectively. PRS derived from the axonogenesis pathway at a *P*-value threshold of $5.5x10^{-2}$ explained 0.017% of the variability (P=$1.6x10^{-2}$) for both left and right UF. PRS derived from the *DCC* gene at a p-value threshold of $7x10^{-2}$ explained 0.05% of the variability (P=$2.5x10^{-5}$; P=$1.3x10^{-4}$) for the left and right UF respectively (Figure 4E). Overall, our results showed that the UF is an important structure contributing to pain and especially

multi-site pain through *DCC*, bridging together for the first-time the genetic determinants of COPC with corticolimbic structures of the human brain.

DISCUSSION

The propensity of chronic pain patients to report more than one location of chronic pain is often observed in clinical settings. Patients diagnosed with one chronic pain condition, such as fibromyalgia, temporomandibular disorder, or headaches, have higher chances of presenting symptoms of other pain conditions[4,5]. Moreover, these patients also report comorbid symptoms such as sleep disturbances, depression, and anxiety[23-25]. Whether COPC is a distinct pathophysiology from the occurrence of single-site chronic pain is unknown[5].

Our analysis of the UK Biobank, one of the largest available datasets, confirmed the high degree of overlap between different chronic pain sites, with one-third of participants with chronic pain reporting multiple pain sites, another third reporting only one pain site, and the remaining third reporting no pain. Our GWAS results showed that distinct genetic factors underlie the report of a single pain condition versus the report of COPC, with multi-site pain having a much stronger genetic component than single-site pain. Furthermore, our study identified a genetic correlation between different chronic pain sites derived from genome-wide data. The strong genetic correlation between chronic pain sites and the causal latent analysis suggests that there is a specific pathway of vulnerability that underlies co-occurring pain conditions, confirming previous observations of twin studies[9]. Headaches, although also highly heritable, did not show genetic overlap with other chronic pain sites, which suggests a distinct pathophysiology. Indeed, previous GWASs of headaches and migraines have shown a strong cardiovascular component[26], whereas in this paper we demonstrated a substantial involvement of CNS components in the genetic pathophysiology of COPCs. Finally, we also confirmed the results of a previous twin study demonstrating a high genetic correlation between widespread pain and abdominal pain[9].

In the field of pain, the majority of existing genetic findings are derived from candidate gene approaches related to specific pain conditions[11,27]. Only recently have large genome-wide studies started to emerge from the UK Biobank for migraine, back pain, as well as multi-site pain, where investigators found many of the SNPs that we uncovered as well[12,13,28]. Here, we aimed to identify the genetic architecture and associated biological pathways of COPC rather than any specific SNP for a specific pain condition and discovered more than 900 variants associated with COPC. These genetic factors explain up to 20% of the variance for multi-site pain, while the heritability for any individual pain site was lower, suggesting a much stronger genetic basis for COPC in comparison with single pain conditions. When we compared the genetic relationship between the report of chronic single-site pain and chronic multi-site pain, we find both common and distinct loci. Contrary to the report of single-site pain, COPC is highly polygenic, with a large portion of its heritability conferred by common genetic variants. The loci that are specific to COPC are enriched in the CNS and are involved in mechanisms related to axonogenesis with a leading role for the *DCC* gene. While the previous studies have found an association between SNPs in *DCC* locus and pain, among many others[12,13] our approaches took single SNP associations results further and identified the central role of *DCC* in the genetics of COPC and uncovered corresponding functional role for netrin and its receptor in the human brain contributing to COPC pathophysiology. Importantly, we also replicated our human findings in another large and independent cohort.

Axon guidance is a process by which neuronal growth cones guide axon extension in the developing nervous system[29]. It involves molecular cues such as netrin 1, present in the environment of growth cones, signaling via dedicated receptors, such as DCC, expressed on the surface of growth cones[14,30-33]. Interestingly, changes in netrin 1 dependent peripheral nerve

outgrowth have been reported in patients with chronic pain[30,34], suggesting that netrin may continue to play an important role following nervous system assembly. The results of the present study further demonstrate that cerebral axonogenesis contributes to COPC. First, heritability partitioning analyses clearly indicated that heritability for multi-site pain was related to genes expressed in the brain. Second, brain imaging data from the Allen Brain Atlas and UK Biobank pointed towards corticolimbic circuits with the UF as a candidate structure for explaining the relationship between the *DCC* gene and COPC.

More specifically, *DCC* gene expression in the human brain appears to be remarkably circumscribed within the basal ganglia and hippocampus. In addition, structural connectivity of the UF was also found to be related to both the *DCC* gene and to multi-site pain. Increased OD values in the UF for multi-site pain suggests that white matter tracts in the UF are less structured in patients exhibiting multi-site pain. This finding seems to be highly consistent with the role of the UF in emotional regulation. The UF, which develops well into the fourth decade of life, connects the medial and lateral orbitofrontal cortex with limbic structures in the temporal lobe such as the amygdala and parahippocampal gyrus[35]. One of the main functions of the UF is to provide subcortical structures with contextual information about potential threats and reward available in the orbitofrontal cortex. As such, UF anatomy has been related to general deficits in the capacity to flexibly predict rewards and punishments, as well as to various neuropsychiatric disorders characterized by emotional dysregulation and poor impulse control, such as major depressive disorder (MDD), attention deficit/hyperactivity disorder (ADHD) and drug abuse[35].

Interestingly, previous studies have shown that the *DCC* gene orchestrates the development of the prefrontal cortex during adolescence[36]. Moreover, GWASs of the UK Biobank have also associated the *DCC* gene with neuropsychiatric disorders characterized by

mood instability such as MDD, post-traumatic stress disorder (PTSD), bipolar disorder (BD), or ADHD[37,38].

Our findings add to these results by linking *DCC* with disorganization of the UF and multi-site pain. Here, we showed that participants who report COPC have higher disorganization in axonal tracks versus participants that report only one pain site or healthy participants. This finding suggests that rewiring of the developing brain predispose to the development of chronic pain. A PRS analysis shed the light on a potential relationship between white matter tract organization in the brain and COPC and showed that variants belonging to *DCC* gene are important mediators of this relationship.

An exclusive involvement of the CNS in pathophysiology of COPC found in our study should be interpreted with caution. Our current results are limited by the broadness of the datasets we use. For instance, our partition heritability analyses did not identify expression from spinal cord, DRGs, or peripheral nerves contributing to multi-site pain. Yet, we are limited here in our analyses of the expression of adult tissues, when we know that *NTN1* and *DCC* are not expressed in the adult spinal cord but only during development. With the increasing broadness of the available expression datasets, new roles for *DCC* may be discovered in addition to that identified here: its crucial contribution to COPC through the wiring of the CNS.

In conclusion, we identified a unique and distinct genetic basis for chronic overlapping pain conditions that points to netrin-driven axonogenesis. Our results suggest that genetically determined *DCC*-dependent axonogenesis in the UF microstructure contributes to COPC via corticolimbic circuits. CNS mechanisms, whether overlapping or distinct, have been suggested as a common neurobiological substrate that may underlie the development of COPC[5,39]. Here, we

identified the genetic and structural basis of this CNS input. Thus, our results suggest a new direction in both fundamental research and therapeutics development.

ONLINE METHODS

Study cohort – UK Biobank

The UK Biobank is a large, prospective, multicenter study of the United Kingdom's population recruited between 2006 and 2010[40,41]. Participants were 40–69 years old and lived within 25 miles of a study recruitment center. Chronic pain conditions were assessed for 502,599 individuals at the initial assessment visit (2006-2010) using a touchscreen-based question: "In the last month, have you experienced any of the following that interfered with your usual activities?" (Data field 6159). The participants had a choice between pain all over the body, back pain, facial pain, headaches, knee pain, stomach/abdominal pain, hip pain, neck/shoulder pain, none of the above and prefer not to answer. For each pain site selected, participants were asked if that pain lasted for more than 3 months (Data fields 2956: pain all over the body; 3404: neck/shoulder pain; 3414: hip pain; 3571: back pain; 3741: stomach/abdominal pain; 3773: knee pain; 3799: headaches; 4067: facial pain). Participants that answered pain all over the body could not indicate any other body site. Cases were defined as individuals self-reporting pain that interfered with their usual activities in the last month and/or that had lasted for more than 3 months. Participants that reported pain at one month and at three months at the same site were defined as having pain chronification. Controls were defined as the participants that answered "none of the above" to data field 6159. Participants that answered, "prefer not to answer" and "do not know" were excluded. Of the 502,599 individuals, 404,381 had phenotype and genotype data available and therefore were analyzed in this paper.

Statistical analysis

Statistical analyses were done using SPSS IBM v 22.0. The prevalence of each chronic pain condition was assessed. The odds ratio (OR) and 95% confidence interval (95% CI) were

calculated to quantify the degree of overlap between conditions. Next, we classified the study population in two groups. The first group included individuals that reported only one pain site that lasted for more than 3 months. The second group included individuals that reported more than one pain site that lasted for more than 3 months, including those who reported widespread pain. This second group was defined as cases reporting multi-site pain as a proxy for chronic overlapping pain conditions (COPC).

Genetic analysis

Out of the 404,381 participants that underwent genotyping and that have available phenotype information, we excluded participants that were not genetically confirmed as "white British", that had sex aneuploidy, or that have a high (≥2%) genotypic missingness rate. After quality control filters were applied, 340,547 participants were considered for analysis. We conducted eight genome-wide association studies (GWASs), one for each pain site, using a logistic regression model to assess heritability and genetic correlations. Next, we also conducted a GWAS contrasting the report of one pain site (n=93,964) with a randomly selected half of participants that answered "none of the above" to data field 6159 (n=81,805). We also conducted a GWAS for chronic multi-site pain, with cases defined as individuals reporting more than one pain site (n=82,812) and controls as the rest of the randomly selected participants that answered "none of the above" to data field 6159 (n=81,966). All genetic analyses were conducted using a logistic regression model with the following co-variates: 40 principle components to account for population stratification, age, age^2, sex, genotyping array, and dummy coded recruitment sites. BOLT-LMM v.2.3 was used in all GWAS analyses, as it accounts for cryptic relatedness[42]. Autosomal analysis was restricted to variants with a MAF >0.1%, info score >0.8, genotype hard call rate >0.95, and Hardy–Weinberg $P > 1\times10^{-12}$. A total of 8,239,177 autosomal makers with

minor allele frequencies above 0.1% that passed quality controls were tested. Heritability was estimated from single nucleotide polymorphisms (SNPs) under an additive model of inheritance using BOLT-REML[42] and LD Score Regression (LDSC)[43].

Genetic correlations were estimated for each pair of pain conditions using LDSC[44]. Tissue-based partitioned heritability was evaluated using LD Score Regression[17,18], with the dataset from the Xavier lab[19].

Gene-based analysis

Gene-based analysis was done using MAGMA. SNPs derived from the summary GWAS were mapped to 18,714 protein-coding genes. A threshold of genome-wide significance level was estimated at $P < 2.67 \times 10^{-6}$.

Genome-wide meta-analysis

In order to identify shared and unique genetic loci between single and multi-site chronic pain summary GWAS datasets, a meta-analysis was performed using GWAMA[16] that was adapted from the sex-specific analysis described previously[45]. The code was adapted to replace the "sex-differentiated" option where we assigned "males" as single-site pain and "females" as multi-site pain[45]. The results of GWAMA will show unique and pleiotropic loci.

Functional mapping and annotation

We used the online platform of FUMA[46] v.1.3.4 to obtain comprehensive annotation information from GWAS summary data. Gene-based tests were obtained using MAGMA[47].

Pathway analyses were conducted with MAGMA within Gene Ontology's (GO) biological processes[48]. Reduction and visualization of GO pathways was done using reviGO[21].

Replication study cohort –HUNT

Participants in the HUNT Study

453 The Nord-Trøndelag Health Study (HUNT) is an ongoing population-based cohort study from
454 the county of Nord-Trøndelag in Norway[49,50]. All inhabitants aged 20 years or older were invited
455 to participate in the HUNT1 survey (1984-1986), the HUNT2 survey (1995-1997), and the
456 HUNT3 survey (2006-2008). Participation rates in HUNT1, HUNT2 and HUNT3 were 89.4%
457 (n=77,212), 69.5% (n=65 237) and 54.1% (n=50 807), respectively[50]. Taken together, the study
458 included more than 120,000 different individuals from Nord-Trøndelag County. For the present
459 study, we included participants from HUNT2 and HUNT3. All participants have provided
460 questionnaire, interview, and measurement data, which can be found at the HUNT databank
461 [https://hunt-db.medisin.ntnu.no/hunt-db]. In addition, about 80,000 participants have provided
462 biological samples for storage at the HUNT biobank [https://www.ntnu.edu/hunt/hunt-biobank].
463 *Phenotype definition in HUNT*
464 The pain questionnaires in HUNT2 and HUNT3 have been described in detail previously[51]. In
465 brief, participants who answered "yes" to the screening question “Have you during the last year
466 continuously for at least 3 months had pain and/or stiffness in muscles and joints?” were
467 requested to indicate the site of the pain, with the possibility to select one or more sites among
468 the following: neck, shoulders, elbows, wrist/hands, upper back, low back, hips, knees, and/or
469 ankles/feet. Cases with chronic multi-site pain were defined as those reporting pain at two or
470 more sites. Controls were defined as those who answered "no" to the screening question on
471 chronic pain. If an individual had participated in both HUNT2 and HUNT3, information from
472 HUNT2 was used. This resulted in a total of 25,747 cases with multi-site pain and 35,753
473 controls without chronic pain.
474
475 *Genotyping, quality control and imputation*

In total, DNA from 71,860 HUNT samples was genotyped using one of three different Illumina HumanCoreExome arrays (HumanCoreExome12 v1.0, HumanCoreExome12 v1.1 and UM HUNT Biobank v1.0). Samples that failed to reach a 99% call rate, had contamination > 2.5% as estimated with BAF Regress[52], large chromosomal copy number variants, lower call rate of a technical duplicate pair and twins, gonosomal constellations other than XX and XY, or whose inferred sex contradicted the reported gender, were excluded. Samples that passed quality control were analyzed in a second round of genotype calling following the Genome Studio quality control protocol described elsewhere[53]. Genomic position, strand orientation and the reference allele of genotyped variants were determined by aligning their probe sequences against the human genome (Genome Reference Consortium Human genome build 37 and revised Cambridge Reference Sequence of the human mitochondrial DNA; http://genome.ucsc.edu) using BLAT[54]. Variants were excluded if their probe sequences could not be perfectly mapped, cluster separation was < 0.3, Gentrain score < 0.15, showed deviations from Hardy Weinberg equilibrium in unrelated samples of European ancestry with p-value < 0.0001), had a call rate < 99%, or another assay with higher call rate genotyped the same variant. Ancestry of all samples was inferred by projecting all genotyped samples into the space of the principal components of the Human Genome Diversity Project (HGDP) reference panel[55,56] (938 unrelated individuals; downloaded from http://csg.sph.umich.edu/chaolong/LASER/), using PLINK. Recent European ancestry was defined as samples that fell into an ellipsoid spanning exclusively European population of the HGDP panel. The different arrays were harmonized by reducing to a set of overlapping variants and excluding variants that showed frequency differences > 15% between data sets, or that were monomorphic in one and had MAF > 1% in another data set. The resulting genotype data were phased using Eagle2 v2.3[57].

Imputation was performed on the 69,715 samples of recent European ancestry using Minimac3[58] (v2.0.1, http://genome.sph.umich.edu/wiki/Minimac3) with default settings (2.5 Mb reference based chunking with 500kb windows) and a customized Haplotype Reference consortium release 1.1 (HRC v1.1) for autosomal variants and HRC v1.1 for chromosome X variants[59]. The customized reference panel represented the merged panel of two reciprocally imputed reference panels: (1) 2,201 low-coverage whole-genome sequences samples from the HUNT study and (2) HRC v1.1 with 1,023 HUNT WGS samples removed before merging. We excluded imputed variants with Rsq < 0.3 or minor allele count <3.

Association testing

We used the Scalable and Accurate Implementation of GEneralized mixed model (SAIGE)[60], which uses a generalized mixed model to account for sample relatedness and cryptic population structure. We ran a mixed logistic regression model, including sex, age, genotyping batch, and the first 4 principal components as covariates. The principal components were calculated by projecting all samples into the space of the principal components of unrelated HUNT samples, using directly genotyped variants in PLINK v1.90[61].

Ethics

The current study is approved by the Regional Committee for Medical and Health Research Ethics (ref. 2015/573).

Allen Brain Atlas

Human gene expression data for visualization of *DCC* expression in the brain were obtained from the Allen Human Brain Atlas (http://human.brain-map.org). A detailed description of this dataset can be found elsewhere[22]. The Neurosynth platform (https://neurosynth.org/) was used extract heat map of normalized expression of *DCC* across the cerebral cortex and subcortical

regions. Visualization of the extracted heat map was done using either Brain Net Viewer[62] or MRICron (https://www.nitrc.org/projects/mricron).

Brain imaging in the UK Biobank

Brain imaging occurred on a subset of subjects at a subsequent brain imaging visit. Inclusion into the pain groups therefore necessitated that subjects met the same chronic pain report on both the initial baseline visit and brain imaging visit. This resulted in 3,985 subjects with no pain, 593 subjects with one-site pain and 800 subjects with multi-site pain. Here, we focused on diffusion-weighted imaging in the UF following the identification of the axonogenesis pathway and the expression of *DCC* in regions the corticolimbic system.

Diffusion data were acquired using a spin-echo echo-planar imaging sequence with two b-values (b = 1,000 and 2,000 s/mm^2) at 2-mm spatial resolution. The diffusion-weighted volumes were acquired with 100 distinct diffusion-encoding directions with multiband acceleration factor of 3. The field of view was 104 × 104 mm, imaging matrix 52 × 52, 72 slices with slice thickness 2 mm, giving 2 mm isotropic voxels. Additional details about the sequence of acquisitions and extraction of IDPs in the UK Biobank can be obtained here: https://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=1977. Briefly, the data was first corrected for eddy currents and head motion using the Eddy tool. Second, the tracts were derived using probabilistic tractography analysis (BEDPOSTx / PROBTRACKx). The automatic mapping of the 27 major white matter tracts was conducted in standard space of each participant using start/stop region of interest masks (implemented using the AutoPtx plugin for FSL). Maps of fractional anisotropy (FA), mean diffusivity (MD), intracellular volume fraction (ICVF), isotropic volume fraction (ISOVF) and orientation dispersion (OD) were registered with the AutoPtx tract masks, allowing the calculation of the averaged value for each parameter across all

voxels pertaining to each tract of interest. Here, we specifically focused on the angular variation in neurite orientation (OD) in the UF.

The OD of neurites can range from highly parallel (coherently oriented white matter structures, such as the corpus callosum) to highly dispersed (gray matter structures characterized by sprawling dendritic processes in all directions).

Polygenic risk scores

Polygenic risk scores (PRSs) were generated using PRSice v.2.3.3[63] using as a base summary GWAS results derived from the single-site and the multi-site GWAS by excluding participants with imaging results. PRSet was used to generate PRSs for the axonogenesis pathway (GO:0007409) and the *DCC* gene with 100 kb on each side. SNPs were clumped using the maximum haplotype frequency estimates and permutation was performed 10,000 times to generate an empirical *P*-values and to prevent Type 1 errors. A regression model that included sex, age, scan site and head scales were used as covariates in a model where each participant's PRS was the dependent variable. A PRS was generated for a series of *P*-value thresholds ($5x10^{-8}$, $1x10^{-7}$, 10^{-6}, 10^{-5}. 10^{-4}, 10^{-3},0.04, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, and 1) in the summary GWAS were to determine the association between pain-related genetic variants and left and right OD of the UF. The best-fit *P*-value threshold was used in the analysis.

ACKNOWLEDGEMENTS

This work was funded by the Canadian Excellence Research Chairs (CERC09). The current study was conducted under UK biobank application no. 20802. The Nord-Trøndelag Health Study (The HUNT Study) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The genotyping was financed by the National Institute of health (NIH), University of Michigan, The Norwegian Research council, and Central Norway Regional Health Authority and the Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU). The genotype quality control and imputation has been conducted by the K.G. Jebsen center for genetic epidemiology, Department of public health and nursing, Faculty of medicine and health sciences, Norwegian University of Science and Technology (NTNU).

CONTRIBUTIONS

SK, MP and LD designed the study and wrote the manuscript. SK and MP performed analyses. The HUNT group provided summary GWAS data for replication. EVP, MR, ST performed the imaging analysis. JM, AK contributed to result interpretation. All the authors read and edited the final manuscript.

HUNT All-In Pain
Anne Heidi Skogholt[1]
Ben Brumpton[1]
Cristen J. Willer[2]
Egil Andreas Fors[3]
Ingrid Heuch[4]
Jonas Bille Nielsen[1,2,5]
Kjersti Storheim[6,7]
Knut Hagen[8]
Kristian Bernhard Nilsen[8,9]

Kristian Hveem[1,10,11]
Lars Fritsche[12]
Laurent F. Thomas[1,13,14,15]
Linda M Pedersen[4]
Maiken E Gabrielsen[1]
Marianne Bakke Johnsen[1,6,16]
Marie Udnesseter Lie[6,16]
Oddgeir Holmen[10]
Sigrid Børte[1,6,16]
Synne Øien Stensland[6,17]
Wei Zhou[18,19]

1 K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
2 Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, 48109, MI, USA.
3 Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
4 Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.
5 Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark.
6 Research and Communication Unit for Musculoskeletal Health (FORMI), Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.
7 Faculty of Health Sciences, Department of physiotherapy, Oslo Metropolitan University, Oslo, Norway.
8 Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.
9 Department of Neurology, Oslo University Hospital, Oslo, Norway.
10 HUNT Research Center, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.
11 Department of Research, Innovation and Education, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.
12 Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, 48109, MI, USA.
13 Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway.
14 BioCore - Bioinformatics Core Facility, Norwegian University of Science and Technology, Trondheim. Norway.
15 Clinic of Laboratory Medicine, St.Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.
16 Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway.
17 NKVTS, Norwegian Centre for Violence and Traumatic Stress Studies.
18 Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA.
19 Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

FIGURE LEGENDS AND TABLES

Figure 1. Pain sites characteristics and correlations in UK Biobank **(A)** Pain sites mapped to the human body. Black dots indicate the sites in the front of the body, while grey dots indicate the sites in the back of the body. Number of cases at each site shown in parenthesis. Human body image from clipart-library.com. **(B)** Epidemiologic and genetic correlations between pain sites. Heatmap showing correlations for co-occurrence of pain sites. Correlations at the epidemiologic odds ratios (OR) are shown in purple hues, while genetic odds ratios (Rg) are shown in orange hues. Grey cells indicate statistical non-significance after Bonferroni correction for the number of same-colored cells. **(C)** Scatterplot showing correlation between epidemiologic OR and body map distance. Each dot is a pair of pain sites out of a total of 21. Also shown are percent variance explained (r^2), slope of regression (m), and associated *P*-value (P). **(D)** Scatterplot showing correlation between genetic Rg and body map distance. **(E)** Scatterplot showing correlation between genetic Rg and epidemiologic OR. **(F)** Narrow-sense heritability estimates for each pain site (blue), for chronic single-site pain (orange), and for chronic multi-site pain (brown). 95% CIs shown in black. The difference in heritability is highly significant (*** $P<2.2\text{x}10^{-16}$).

Figure 2. Genome-wide association studies for single-site pain and multi-site pain. Shown are Manhattan plots at the SNP-level (top) and at the gene-level (bottom). SNP *P*-values are obtained from BOLT or GWAMA, while gene *P*-values are obtained from MAGMA. Alternating dark and light color hues used for odd and even chromosome numbers. Genome-wide significance highlighted by a horizontal red line at SNP-level is from Bonferroni's threshold of $5\text{x}10^{-8}$, while at gene-level is at FDR 1%. **(A)** Single-vs-no chronic pain site. **(B)** Multi-site-vs-no chronic pain sites. **(C)** Unique loci derived from a meta-analysis in GWAMA. **(D)** Pleiotropic loci from a meta-analysis in GWAMA.

Figure 3. Partitioned heritability for single-site pain and multi-site pain. **(A)** Seventy-eight tissues were grouped into eight tissue classes: central nervous system (CNS, green, *n*=21), peripheral nervous system (PNS, blue, *n*=4), endocrine (END, purple, *n*=2), myeloid (MYE, red, *n*=16), B cells (B, orange, *n*=8) T cells (T, purple, *n*=22), adipose (ADI, brown, *n*=2) and muscle (MUS, grey, *n*=3). Shown for each tissue is –log10 of FDR-adjusted *P*-value for enrichment. Heritability estimated for single-site pain (top) and for multi-pain sites are shown (COPC; bottom). Statistical threshold of significance is highlighted at the FDR 10% level with horizontal red lines, while significant tissues with colored filled boxes. **(B)** Zoom into the CNS tissues for multi-site pain. **(C)** Scatter plot of heritability coefficients in single-site pain versus multi-site pain. Each dot is a tissue of the CNS. Orange line obtained from linear regression, with percent variance explained (r^2), slope (m) and regression *P*-value (P) shown.

Figure 4. Functional validation for a role of *DCC* in the human brain. **(A)** Whole brain expression of *DCC* computed from the Allen Brain Atlas. **(B)** Zoom into the expression of *DCC* in the subcortical limbic regions. **(C)** Representation of the uncinate fasciculus (UF) white matter tract. **(D)** Bar plot of bilateral dispersion orientation (OD) of the UF in the no-pain controls, single-site pain, multi-site pain states. The *y*-axis represents OD values for the UF. Bars represent standard error. **P*<0.05; ****P*<0.0001. **(E)** Polygenic risk score (PRS) generated using PRSice from summary GWAS of single-site pain, multi-site pain, axonogenesis pathway, and *DCC*. Plotted is the -log 10 *P*-value of the regression model using PRS with the score selected at the best fit *P*-value threshold.

Table 1. Demographic and phenotypic characteristics of the study population.

Table 2. Replication of results on multi-site pain from UK Biobank in HUNT.

SUPPLEMENTARY MATERIALS

Supplementary Figure 1. Histogram of number of UK Biobank participants per reported number of chronic pain sites.

Supplementary Figure 2. QQ plot: Quantile-quantile plot shows the observed versus expected – log10 p-values from A) one pain site and B) multi-site pain association analysis.

Supplementary Figure 3. Locus Zoom plots for each of the 23 genome-wide significant loci.

Supplementary Table 1. Prevalence of acute and chronic pain sites in UK Biobank.

Supplementary Table 2. Epidemiological odds of reporting pairs of chronic pain sites.

Supplementary Table 3. Genetic correlation between pairs of chronic pain sites.

Supplementary Table 4. Latent causal variable analysis between chronic pain sites.

Supplementary Table 5. Heritability estimates for chronic pain sites.

Supplementary Table 6. List of top SNPs associated with single-site pain.

Supplementary Table 7. List of protein-coding genes associated with single-site pain.

Supplementary Table 8. List of genome-wide loci associated with multi-site pain.

Supplementary Table 9. List of protein-coding genes associated with multi-site pain.

Supplementary Table 10. List of protein-coding genes derived from GWAMA that are unique for single-site pain or multi-site pain GWASs.

Supplementary Table 11. List of protein-coding genes derived from GWAMA that are pleiotropic between single-site pain or multi-site pain GWASs.

Supplementary Table 12. Tissue-specific partitioned heritability within the Xavier lab dataset for single-site pain.

Supplementary Table 13. Tissue-specific partitioned heritability within the Xavier lab dataset for multi-site pain.

706 **Supplementary Table 14.** Pathway-based functional analyses for single-site pain GWAS. **(A)**
707 Analysis in Gene Ontology's biological processes.
708 **Supplementary Table 15.** Pathway-based functional analyses for multi-site pain GWAS. **(A)**
709 Analysis in Gene Ontology's biological processes. **(B)** Reduced pathway sets from reviGO.
710 **Supplementary Table 16.** Replication in HUNT cohort. **(A)** Locus-level; **(B)** Gene-level; **(C)**
711 Pathway level.
712 **Supplementary Table 17.** Polygenic risk score (PRS) regression models testing left, right, and
713 bilateral orientation dispersion of the uncinate fasciculus (UF).
714

REFERENCES

1. Fayaz, A., Croft, P., Langford, R.M., Donaldson, L.J. & Jones, G.T. Prevalence of chronic pain in the UK: a systematic review and meta-analysis of population studies. *BMJ Open* **6**, e010364 (2016).
2. Shupler, M.S., Kramer, J.K., Cragg, J.J., Jutzeler, C.R. & Whitehurst, D.G.T. Pan-Canadian estimates of chronic pain prevalence from 2000 to 2014: A Repeated Cross-Sectional Survey Analysis. *J Pain* (2018).
3. Schopflocher, D., Taenzer, P. & Jovey, R. The prevalence of chronic pain in Canada. *Pain Res Manag* **16**, 445-50 (2011).
4. Page, M.G., Fortier, M., Ware, M.A. & Choiniere, M. As if one pain problem was not enough: prevalence and patterns of coexisting chronic pain conditions and their impact on treatment outcomes. *J Pain Res* **11**, 237-254 (2018).
5. Maixner, W., Fillingim, R.B., Williams, D.A., Smith, S.B. & Slade, G.D. Overlapping Chronic Pain Conditions: Implications for Diagnosis and Classification. *J Pain* **17**, T93-t107 (2016).
6. Larkin, T.E. *et al.* Altered network architecture of functional brain communities in chronic nociplastic pain. *Neuroimage* **226**, 117504 (2021).
7. Schrepf, A. *et al.* ICD-10 Codes for the Study of Chronic Overlapping Pain Conditions in Administrative Databases. *J Pain* **21**, 59-70 (2020).
8. Nielsen, C.S., Knudsen, G.P. & Steingrimsdottir, O.A. Twin studies of pain. *Clin Genet* **82**, 331-40 (2012).
9. Vehof, J., Zavos, H.M., Lachance, G., Hammond, C.J. & Williams, F.M. Shared genetic factors underlie chronic pain syndromes. *Pain* **155**, 1562-8 (2014).
10. Gasperi, M. *et al.* Chronic prostatitis and comorbid non-urological overlapping pain conditions: A co-twin control study. *J Psychosom Res* **102**, 29-33 (2017).
11. Meloto, C.B. *et al.* Human pain genetics database: a resource dedicated to human pain genetics research. *Pain* **159**, 749-763 (2018).
12. Suri, P. *et al.* Genome-wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. *PLoS Genet* **14**, e1007601 (2018).
13. Johnston, K.J.A. *et al.* Genome-wide association study of multisite chronic pain in UK Biobank. *PLoS Genet* **15**, e1008164 (2019).
14. Kennedy, T.E., Serafini, T., de la Torre, J.R. & Tessier-Lavigne, M. Netrins are diffusible chemotropic factors for commissural axons in the embryonic spinal cord. *Cell* **78**, 425-35 (1994).
15. Yamagata, M., Weiner, J.A. & Sanes, J.R. Sidekicks: synaptic adhesion molecules that promote lamina-specific connectivity in the retina. *Cell* **110**, 649-60 (2002).
16. Mägi, R. & Morris, A.P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
17. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
18. Finucane, H.K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629 (2018).
19. Benita, Y. *et al.* Gene enrichment profiles reveal T-cell development, differentiation, and lineage-specific transcription factors including ZBTB25 as a novel NF-AT repressor. *Blood* **115**, 5376-84 (2010).
20. Fehrmann, R.S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47**, 115-25 (2015).
21. Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
22. Hawrylycz, M.J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399 (2012).

23. Aaron, L.A., Burke, M.M. & Buchwald, D. Overlapping conditions among patients with chronic fatigue syndrome, fibromyalgia, and temporomandibular disorder. *Arch Intern Med* **160**, 221-7 (2000).
24. Nickel, J.C. *et al.* Interstitial cystitis/painful bladder syndrome and associated medical conditions with an emphasis on irritable bowel syndrome, fibromyalgia and chronic fatigue syndrome. *J Urol* **184**, 1358-63 (2010).
25. Schafer, I. *et al.* Reducing complexity: a visualisation of multimorbidity by combining disease clusters and triads. *BMC Public Health* **14**, 1285 (2014).
26. Gormley, P. *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat Genet* **48**, 856-66 (2016).
27. Zorina-Lichtenwalter, K., Meloto, C.B., Khoury, S. & Diatchenko, L. Genetic predictors of human chronic pain conditions. *Neuroscience* **338**, 36-62 (2016).
28. Meng, W. *et al.* A Genome-Wide Association Study Finds Genetic Associations with Broadly-Defined Headache in UK Biobank (N=223,773). *EBioMedicine* **28**, 180-186 (2018).
29. Glasgow, S.D. *et al.* Activity-Dependent Netrin-1 Secretion Drives Synaptic Insertion of GluA1-Containing AMPA Receptors in the Hippocampus. *Cell Rep* **25**, 168-182 e6 (2018).
30. Wu, C.H. *et al.* Netrin-1 Contributes to Myelinated Afferent Fiber Sprouting and Neuropathic Pain. *Mol Neurobiol* **53**, 5640-51 (2016).
31. Serafini, T. *et al.* The netrins define a family of axon outgrowth-promoting proteins homologous to C. elegans UNC-6. *Cell* **78**, 409-24 (1994).
32. Keino-Masu, K. *et al.* Deleted in Colorectal Cancer (DCC) encodes a netrin receptor. *Cell* **87**, 175-85 (1996).
33. da Silva, R.V. *et al.* DCC Is Required for the Development of Nociceptive Topognosis in Mice and Humans. *Cell Rep* **22**, 1105-1114 (2018).
34. Schubert, A.L., Held, M., Sommer, C. & Üçeyler, N. Reduced gene expression of netrin family members in skin and sural nerve specimens of patients with painful peripheral neuropathies. *J Neurol* **266**, 2812-2820 (2019).
35. Olson, I.R., Von Der Heide, R.J., Alm, K.H. & Vyas, G. Development of the uncinate fasciculus: Implications for theory and developmental disorders. *Dev Cogn Neurosci* **14**, 50-61 (2015).
36. Manitt, C. *et al.* dcc orchestrates the development of the prefrontal cortex during adolescence and is altered in psychiatric patients. *Transl Psychiatry* **3**, e338 (2013).
37. Wu, Y. *et al.* Multi-trait analysis for genome-wide association study of five psychiatric disorders. *Transl Psychiatry* **10**, 209 (2020).
38. Ward, J. *et al.* Genome-wide analysis in UK Biobank identifies four loci associated with mood instability and genetic correlation with major depressive disorder, anxiety disorder and schizophrenia. *Transl Psychiatry* **7**, 1264 (2017).
39. Yang, Q., Wang, Z., Yang, L., Xu, Y. & Chen, L.M. Cortical thickness and functional connectivity abnormality in chronic headache and low back pain patients. *Hum Brain Mapp* **38**, 1815-1832 (2017).
40. Allen, N.E., Sudlow, C., Peakman, T., Collins, R. & Biobank, U.K. UK biobank data: come and get it. *Sci Transl Med* **6**, 224ed4 (2014).
41. Sudlow, C. *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* **12**, e1001779 (2015).
42. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-90 (2015).
43. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
44. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
45. Magi, R., Lindgren, C.M. & Morris, A.P. Meta-analysis of sex-specific genome-wide association studies. *Genet Epidemiol* **34**, 846-53 (2010).

46. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).
47. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).
48. The_Gene_Ontology_Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**, D330-D338 (2019).
49. Holmen, T.L. *et al.* Cohort profile of the Young-HUNT Study, Norway: a population-based study of adolescents. *Int J Epidemiol* **43**, 536-44 (2014).
50. Krokstad, S. *et al.* Cohort Profile: the HUNT Study, Norway. *Int J Epidemiol* **42**, 968-77 (2013).
51. Hagen, K., Linde, M., Heuch, I., Stovner, L.J. & Zwart, J.A. Increasing prevalence of chronic musculoskeletal complaints. A large 11-year follow-up in the general population (HUNT 2 and 3). *Pain Med* **12**, 1657-66 (2011).
52. Jun, G. *et al.* Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* **91**, 839-48 (2012).
53. Guo, Y. *et al.* Illumina human exome genotyping array clustering and quality control. *Nat Protoc* **9**, 2643-62 (2014).
54. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
55. Wang, C. *et al.* Ancestry estimation and control of population stratification for sequence-based association studies. *Nat Genet* **46**, 409-15 (2014).
56. Li, J.Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100-4 (2008).
57. Loh, P.R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448 (2016).
58. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
59. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
60. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335-1341 (2018).
61. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
62. Xian, M., Wang, J. & He, Y. BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *Plos One* **8**, e68910 (2013).
63. Choi, S.W. & O'Reilly, P.F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* **8**(2019).

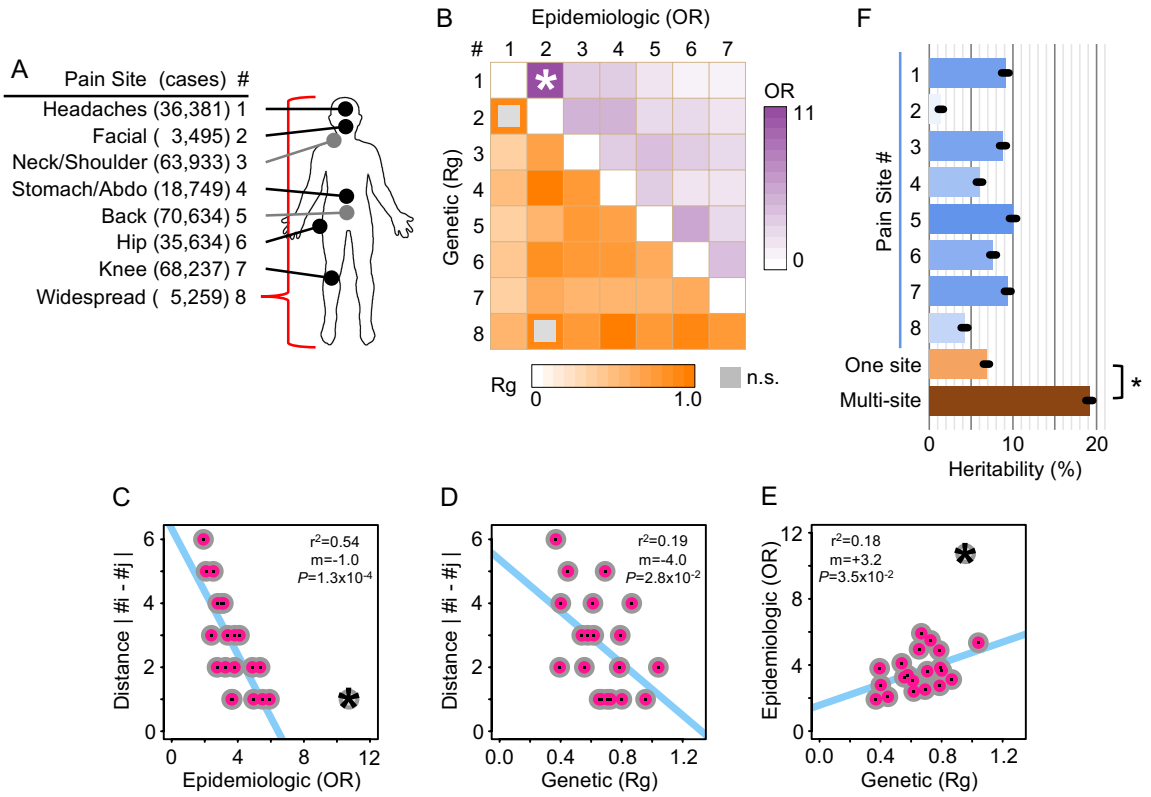

Figure 1. Pain sites characteristics and correlations in UK biobank **(A)** Pain sites mapped to the human body. Black dots indicate the sites in the front of the body, while grey dots indicate the sites in the back of the body. Number of cases at each site shown in parenthesis. Human body image from clipart-library.com. **(B)** Epidemiologic and genetic correlations between pain sites. Heatmap showing correlations for co-occurrence of pain sites. Correlations at the epidemiologic odds ratios (OR) are shown in purple hues, while genetic odds ratios (Rg) are shown in orange hues. Grey cells indicate statistical non-significance after Bonferroni correction for the number of same-colored cells. **(C)** Scatter plot showing correlation between epidemiologic OR and body map distance. Each dot is a pair of pain sites out of a total of 21. Also shown are percent variance explained (r2), slope of regression (m), and associated P-value (P). **(D)** Scatter plot showing correlation between genetic Rg and body map distance. **(E)** Scatter plot showing correlation between genetic Rg and epidemiologic OR. **(F)** Narrow-sense heritability estimates for each pain site (blue), for chronic single-site pain (orange), and for chronic multi-site pain (brown). 95% confidence intervals shown in black. The difference in heritability is highly significant (*** P<2.2x10-16).

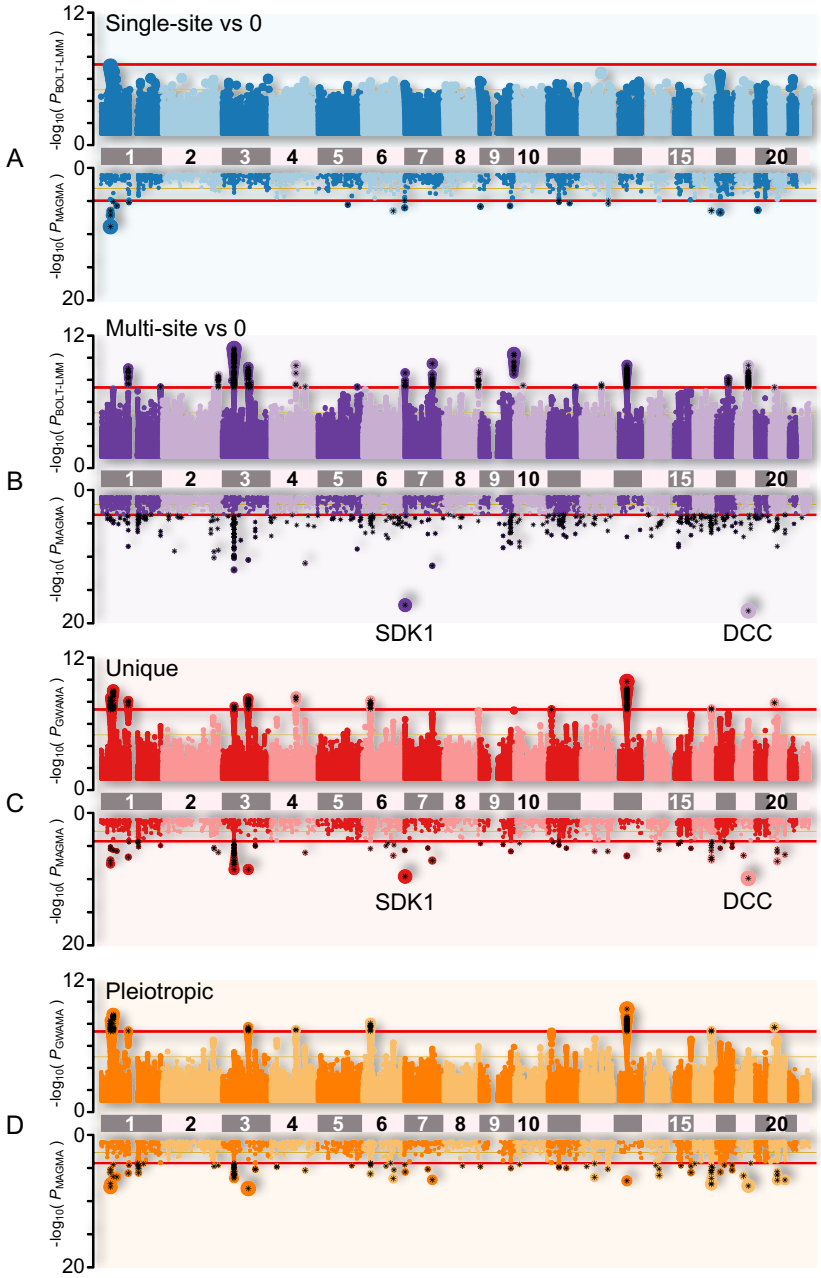

Figure 2. Genome-wide association studies for single-site pain and multi-site pain. Shown are Manhattan plots at the SNP-level (top) and at the gene-level (bottom). SNP P-values are obtained from BOLT or GWAMA, while gene P-values are obtained from MAGMA. Alternating dark and light color hues used for odd and even chromosome numbers. Genome-wide significance highlighted by a horizontal red line at SNP-level is from Bonferroni's threshold of 5x10-8, while at gene-level is at FDR 1%. **(A)** Single-site-vs-no chronic pain site. **(B)** Multi-site-vs-no chronic pain sites. **(C)** Unique loci derived from a meta-analysis in GWAMA. **(D)** Pleiotropic loci from a meta-analysis in GWAMA.

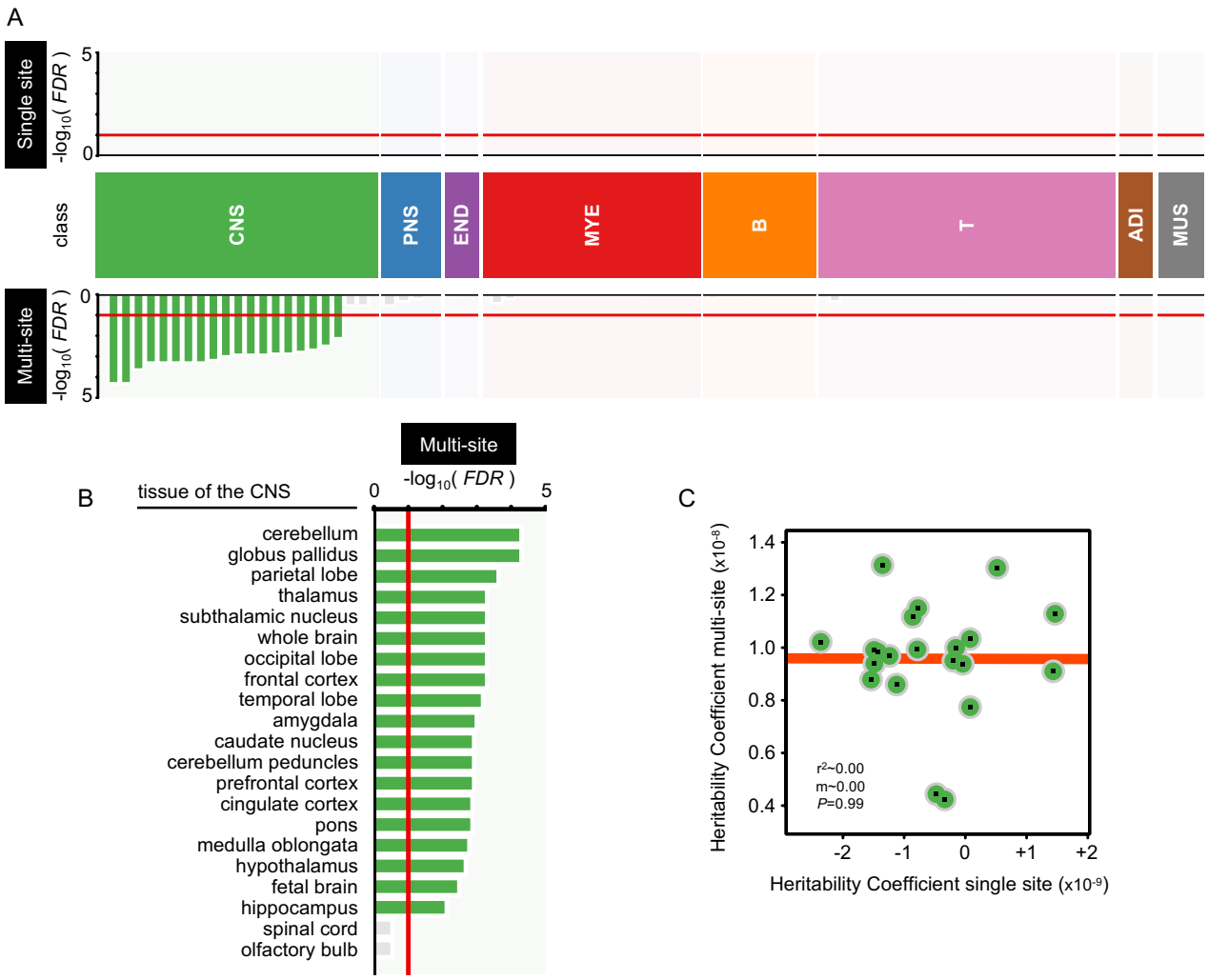

Figure 3. Partitioned heritability for single-site pain and multi-site pain. **(A)** Seventy-eight tissues were grouped into eight tissue classes: central nervous system (CNS, green, n=21), peripheral nervous system (PNS, blue, n=4), endocrine (END, purple, n=2), myeloid (MYE, red, n=16), B cells (B, orange, n=8) T cells (T, purple, n=22), adipose (ADI, brown, n=2) and muscle (MUS, grey, n=3). Shown for each tissue is –log10 of FDR-adjusted P-value for enrichment. Heritability estimated for single-site pain (top) and for multi-pain sites are shown (COPC; bottom). Statistical threshold of significance is highlighted at the FDR 10% level with horizontal red lines, while significant tissues with colored filled boxes. **(B)** Zoom into the central nervous system tissues for multi-site pain. **(C)** Scatter plot of heritability coefficients in single-site pain versus multi-site pain. Each dot is a tissue of the CNS. Orange line obtained from linear regression, with percent variance explained (r2), slope (m) and regression P-value (P) shown.

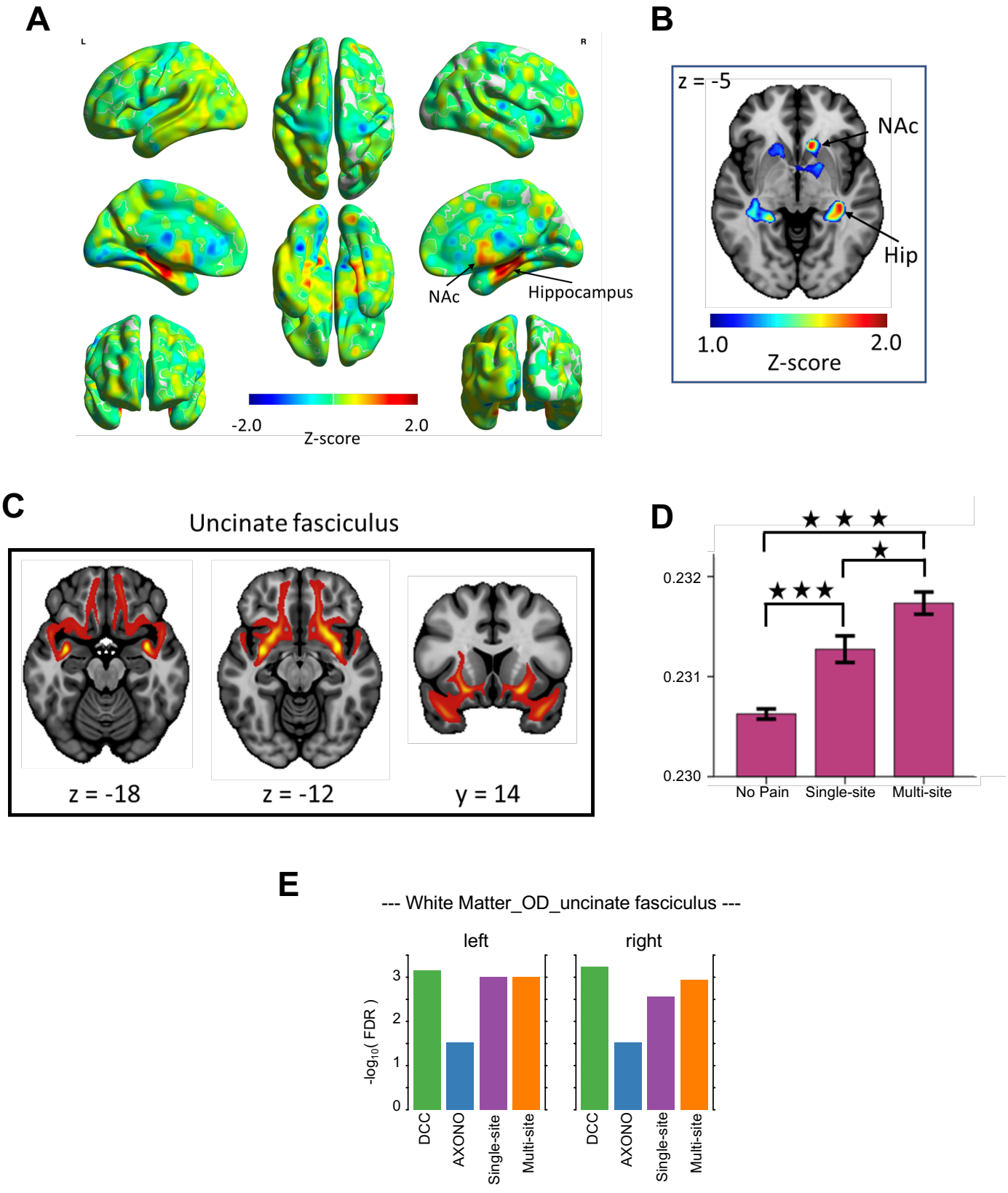

Figure 4. Functional validation for a role of DCC in the human brain (A) Whole brain expression of DCC computed from the Allen Brain Atlas (B) Zoom into the expression of DCC in the subcortical limbic regions (C) Representation of the uncinate fasciculus (UF) white matter tract (D) Bar plot of bilateral dispersion orientation (OD) of the UF in the no-pain controls, single-site pain, multi-site pain states. The Y-axis represents OD values for the UF. Bars represent standard error. ***p<0.0001; *p=0.02 (E) Polygenic risk score generated using PRSice from summary GWAS of single-site pain, multi-site pain, axonogenesis pathway, and DCC. Plotted is the -log 10 p-value of the regression model using PRS with the score selected at the best fit p-value threshold.

Table 1 – Demographic and phenotypic characteristics of study population

	Controls	One-site	Multi-site	P-value
Females (%)	52.4%	54.2%	60.7%	<0.0001
Age (mean)	56.78	56.67	56.98	<0.0001
BMI (mean)	26.70	27.67	28.66	<0.0001
Smoking status (current)	8.8%	10.8%	13.6%	<0.0001
Townsend deprivation index (mean)	-1.60	-1.32	-0.80	<0.0001
Number of self-reported cancers	0.09	0.09	0.1	<0.0001
Number of self-reported non-cancer illnesses	1.44	1.94	2.83	<0.0001
Medication for pain relief				
Paracetamol (n)	20,846	28,800	40,954	<0.0001
Ibuprofen (n)	14,480	21,137	24,468	<0.0001
Aspirin (n)	23,418	16,278	17,602	<0.0001
Depressed mood last two weeks				<0.0001
Severe days	12.9%	18.9%	25.6%	
More than half the days	1.6%	3.0%	5.5%	
Nearly every day	0.9%	1.7%	4.4%	
Number of depression episodes (mean)	2.44	2.78	3.21	<0.0001
Neuroticism score (mean)	3.35	4.32	5.41	<0.0001

Categorical data were compared using a chi-square test and quantitative data are compared using a t-test. The overall p-value is an ANOVA between the three groups.

Table 2 – Replication of multi-site chronic pain results from UK biobank in HUNT

a) Loci SNP level

Loci	Lead SNP	Genes in locus	HUNT p-value
Chr3 :49,206,000-49,891,000	rs11709734	*APEH, BSN, C3orf62, C3orf84, CCDC36, CDHR4, DGA1, GMPPB, GPX1, IP6K1, KLHDC81B, MST1, MST1R, NICN1, RHOA, RNF123, TCTA, TRAIP, UBA7, USP4*	rs184219667 (r^2=0.96) $1.36x10^{-3}$
Chr4 :140,600,000-141,000,000	rs34595097	*MAML3*	rs1204594 (r^2=0.54) $2.33x10^{-4}$
Ch7 :113,770,000-114,267,000	rs12672683	*FOXP2*	rs62469212 (r^2=0.51) $1.42x10^{-3}$
Chr18 : 50,073,000-50,908,000	rs8099145	*DCC*	rs17410557 (r^2=0.58) $1.68x10^{-4}$

b) Gene level

HUGO	CHR	START	STOP	Z stat	HUNT P-value	FDR	
DCC	18	49866542	51062273	5.44	2.64E-08	0.000497	DCC netrin 1 receptor
CAMKV	3	49895414	49907655	4.10	2.00E-05	0.047772	CaM Kinase like vesicle associated
IP6K1	3	49761728	49823973	4.10	2.03E-05	0.047772	Inositol hexakisphosphate kinase 1
MON1A	3	49946302	49967445	4.01	3.09E-05	0.058253	MON1 homolog A, secretory trafficking associated
MAML3	4	1.41E+08	1.41E+08	3.90	4.82E-05	0.070071	Mastermind loke transcriptional coactivator 3
RNF123	3	49726950	49758962	3.68	0.000119	0.083047	Ring finger protein 123
ZBTB46	20	62375021	62463731	3.53	0.000209	0.108017	Zinc finger and BTB Domain containing 46
BSN	3	49591922	49708982	3.45	0.000284	0.118353	Bassoon presynaptic cytomatrix protein
TRAIP	3	49866028	49893992	3.38	0.000357	0.126412	TRAF interacting protein
RBM6	3	49977474	50114685	3.32	0.000454	0.144758	RNA binding motif protein 6
MST1	3	49721380	49726196	3.26	0.00056	0.159801	Macrophage stimulating 1

c) Pathway level

VARIABLE	DESC	HUNT P-value	FDR
GO:0007409	axonogenesis	0.00095495	0.547171
GO:0061564	axon development	0.0013778	0.547171
GO:0042297	vocal learning	0.0014606	0.547171
GO:0098596	imitative learning	0.0014606	0.547171
GO:0048812	neuron projection morphogenesis	0.0023204	0.595695
GO:0048667	cell morphogenesis involved in neuron differentiation	0.0023438	0.595695
GO:0120039	plasma membrane bounded cell projection morphogenesis	0.0024645	0.595695
GO:0048858	cell projection morphogenesis	0.0026534	0.595695
GO:0006206	pyrimidine nucleobase metabolic process	0.0036173	0.612315
GO:0098597	observational learning	0.0040134	0.612315
GO:0032913	negative regulation of transforming growth factor beta3 production	0.0040365	0.612315
GO:0007638	mechanosensory behavior	0.0058522	0.621343
GO:0032990	cell part morphogenesis	0.0058527	0.621343
GO:0098598	learned vocalization behavior or vocal learning	0.0059061	0.621343
GO:0010608	posttranscriptional regulation of gene expression	0.0069535	0.63441
GO:0031223	auditory behavior	0.0069601	0.63441
GO:0030182	neuron differentiation	0.0099669	0.690212
GO:0007399	nervous system development	0.03211	0.773552
GO:0022008	neurogenesis	0.033344	0.773786
GO:0071625	vocalization behavior	0.03577	0.778559
GO:0048468	cell development	0.038141	0.78672
GO:0010468	regulation of gene expression	0.042054	0.78672
GO:0000904	cell morphogenesis involved in differentiation	0.044285	0.78672
GO:0006208	pyrimidine nucleobase catabolic process	0.047884	0.78672