

## Genome-wide analysis identifies significant contribution of brain-expressed genes in chronic, but not acute, back pain

Andrey V. Bortsov<sup>1,†</sup>, Marc Parisien<sup>2,†</sup>, Samar Khoury<sup>2,†</sup>, Dmitri V. Zaykin<sup>3</sup>, Amy E Martinsen<sup>5,6,7</sup>, Marie Udnesseter Lie<sup>4,5,6</sup>, Ingrid Heuch<sup>7</sup>, *HUNT All-In Pain*, Kristian Hveem<sup>5,8</sup>, John-Anker Zwart<sup>5,6,7</sup>, Bendik S Winsvold<sup>5,7</sup>, Luda Diatchenko<sup>2\*</sup>

1. *Center for Translational Pain Medicine, Department of Anesthesiology, Duke University, Durham, NC, USA.*

2. *Alan Edwards Centre for Research on Pain; Department of Anesthesia, School of Medicine; School of Dentistry; McGill University, Montréal, QC, Canada.*

3. *Biostatistics and Computational Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA.*

4. *Research and Communication Unit for Musculoskeletal Health (FORMI), Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.*

5. *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.*

6. *Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway.*

7. *Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway.*

8. *HUNT Research Center, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway.*

† These authors equally contributed to this work

\* Corresponding author: [luda.diatchenko@mcgill.ca](mailto:luda.diatchenko@mcgill.ca)

This manuscript has 4 figures, 2 tables, and 2 supplementary tables.

## **ABSTRACT**

Back pain is the leading cause of disability worldwide. Although most cases of back pain are acute, 20% of people with acute back pain go on to experience chronic back pain symptoms. It is unclear if acute and chronic pain states have similar or distinct underlying genetic mechanisms. Here we performed a genome-wide analysis for acute and chronic back pain in 375,158 individuals and found a significant genetic contribution to chronic, but not to acute, back pain. Using the UK Biobank cohort for discovery and the HUNT cohort for replication, we identified 7 loci for chronic back pain, of which 3 are novel. Pathway analyses, tissue-specific heritability enrichment analyses, epigenetic characterization, and tissue-specific transcriptome mapping in mouse pain models suggest a substantial genetic contribution to chronic, but not acute, back pain from the loci predominantly expressed in the central nervous system. Our findings show that chronic back pain is more heritable than acute back pain and is driven mostly by genes expressed in the central nervous system.

## INTRODUCTION

Back pain is the world's leading cause of disability, reducing quality of life, and imposing significant health care costs.<sup>1-4</sup> According to the International Association for the Study of Pain (IASP) criteria, a back pain episode lasting less than 3 months is defined as acute back pain while pain lasting for 3 months or longer is defined as chronic.<sup>5</sup> Although the majority of individuals with acute back pain resolve their pain,<sup>6,7</sup> a substantial proportion of these cases transition to chronic back pain. As a result, chronic back pain is the most common painful condition which affects 10-15% of the adult population.<sup>2,3,8</sup> Mechanisms involved in the development of chronic back pain remain largely unknown, and seem to be different from causes leading to the onset of acute back pain. While some believe that the majority (up to 85%) of cases of chronic back pain cannot be attributed to a known specific pathology,<sup>9</sup> others attribute chronic back pain to processes such as central sensitization,<sup>10</sup> disc herniation causing inflammation of the affected nerve,<sup>11</sup> facet joint syndrome,<sup>12</sup> sacroiliac joint pain,<sup>13-15</sup> lumbar spinal stenosis<sup>16</sup>, and lumbar disc degeneration.<sup>17</sup>

Heritability estimates for back pain from twin studies range from 30 to 68%.<sup>18-21</sup> Twin studies also show shared heritability between chronic back pain and lumbar disc degeneration,<sup>18</sup> depression,<sup>22</sup> anxiety,<sup>22</sup> education,<sup>23</sup> obesity,<sup>24</sup> chronic widespread musculoskeletal pain.<sup>25</sup> Recent genome-wide association studies (GWAS) using the UK Biobank cohort have identified a number of genes associated with chronic back pain, including *SPOCK2*, *SOX5*, and *DCC*.<sup>26,27</sup> These studies confirm genetic involvement both in the etiology of chronic back pain and its related comorbidities.

Despite the mounting evidence that acute and chronic back pain are distinct states with different pathologies, there is a significant gap in knowledge about the genetic risk factors and

molecular pathophysiology associated with acute versus chronic back pain. Previous genetic studies have looked either at back pain in general,<sup>26</sup> or only at the chronic pain condition.<sup>27</sup> One important question is whether acute and chronic pain states have similar underlying genetic pathways. Here we used GWAS for both acute and chronic back pain states and characterized the corresponding molecular and cellular pathways contributing to these genetics associations. We hypothesized that acute and chronic back pain were different with respect to individual genetic risk loci, overall genetic heritability, heritability distribution across tissues and cells, and contributing molecular and cellular pathways.

## RESULTS

In our genome-wide exploration of genetic factors contributing to acute versus chronic back pain we used data on individuals of white British ancestry from the UK Biobank cohort (n=375,158) (Fig. 1). Acute and chronic back pain phenotypes were inferred from an affirmative response to the question of whether the respondent had experienced back pain in the last month that interfered with their usual activities. Those who reported back pain in the last month were asked if their back pain lasted for more than 3 months. Acute back pain was defined as back pain (that interfered with activities) in the last month that lasted no more than three months. Chronic back pain was defined as back pain that lasted for more than three months. Individuals with no acute or chronic back pain (but possibly with other types of pain) were used as controls. The defined phenotypes were used as outcome variables to perform the GWAS.

A genome-wide association analysis of chronic back pain (n=375,158; 70,633 cases; 304,525 controls) revealed 355 SNPs that reached the genome-wide significance threshold ( $5 \times 10^{-8}$ ), grouped into 13 loci (Fig. 1A, Supplementary Table 1). The majority of the significant

loci (8 of 13) have already been reported as significantly associated with back pain in the UK Biobank cohort by other groups (chr4: TMED11P • SPON2, chr8: C8orf34-AS1 • C8orf34 • LINC01592, chr8: CCDC26 • GSDMC, chr9: NRARP • EXD3 • NOXA1, chr10: PSAP • CHST3 • SPOCK2 • ASCC1, chr10: BNIP3 • JAKMIP3 • DPYSL4, chr12: LOC101928441 • SOX5 • SOX5-AS1, chr18: LINC01630 • DCC • LINC01919), and five loci were new (chr5: LOC105374704 • CDH6, chr5: NUDT12 • RAB9BP1, chr10: LINC00844 • CCEPR, chr14: SLC25A21-AS1 • MIPOL1 • FOXA1, chr19: BCAM • NECTIN2 • TOMM40 • APOE • APOC1 • APOC1P1).<sup>26,27</sup> The estimated narrow-sense SNP heritability ( $h^2$ ) for chronic back pain was 4.6% ( $P=1.4 \times 10^{-78}$ , Fig. 1A). While there was some inflation of test statistics (genomic control parameter  $\lambda_{GC} = 1.31$ ), the LD Score regression intercept was close to 1 (Fig. 1B), indicating that the inflation was due to polygenicity rather than confounding factors such as population stratification and/or cryptic relatedness, the use of a mixed model, or population stratification.<sup>28</sup> The latter confounding factors were mitigated using BOLT-LMM via genetic principal components and kinship covariance matrices.<sup>29,30</sup>

The GWAS of acute back pain ( $n=336,734$ ; 32,209 cases and 304,525 controls) revealed no genetic variants reached the genome-wide statistical significance threshold (Fig. 1C). The estimated narrow-sense SNP heritability ( $h^2=0.81\%$ ) for acute back pain was also lower than for chronic back pain; however, the heritability was significantly different from zero ( $P=1.4 \times 10^{-8}$ , Fig. 1C). The LD Score regression intercept was 1, suggesting that the moderate inflation ( $\lambda_{GC} = 1.05$ ) was due to polygenicity rather than confounding factors (Fig. 1D). Of note, amongst the 13 statistically significant loci in chronic back pain, only one locus on chromosome 8 was associated with both acute and chronic back pain (Supplementary Table 2).

In the HUNT replication cohort, we found similar differences in heritability estimates between acute and chronic back pain. Heritability of chronic back pain was 3.4% ( $P=0.0011$ ), whereas heritability of acute back pain was 0.6% ( $P=0.851$ ). From this we concluded that the heritability of acute back pain is approximately 5-6 times smaller than the heritability of chronic back pain. A total of 7 out of 13 statistically significant loci were replicated in the HUNT cohort for chronic pain cases (Table 1) at the  $P < 0.05$  level using sets of SNPs at high LD ( $r > 0.8$ ).

At the pathway level, a total of 21 Gene Ontology (GO) pathways were significantly enriched (FDR 20%) in chronic back pain and 10 pathways were enriched (FDR 20%) in acute back pain in the UK Biobank discovery cohort (Table 2). Importantly, different pathways were identified for acute and chronic pain states. Neurogenesis and synaptic plasticity were significant in chronic back pain while odontogenesis, cardiac muscle depolarization, and immune response through Th2-helpers were significant in acute back pain. Of note, the significance of the odontogenesis pathway is driven by genes linked to connective tissue disorders (*RSPO2*) and bone remodeling (*TNFRSF11B*).<sup>31,32 33</sup> Three pathways for chronic pain were replicated ( $P < 0.05$ ) in the HUNT cohort: Spinal cord ventral commissure morphogenesis (GO:0021965), Positive regulation of kinase activity (GO:0033674,) and Positive regulation of transferase activity (GO:0051347). None of the acute pain pathways were replicated in the HUNT cohort.

We next investigated how the genetic heritabilities of chronic and acute back pain were distributed across different tissues. We measured the enrichment for the genes corresponding to identified genomic loci expressed in various human tissues. A total of 152 cell types and tissues from Fehrmann *et al*,<sup>34</sup> grouped into eight categories (adipose, blood and immune, cardiovascular, central nervous system (CNS), digestive, endocrine, musculoskeletal and connective, and other tissues) were included in the analyses (Fig. 2). For chronic back pain,

estimates for heritability that exhibited the smallest  $P$ -values were attributed to genomic regions expressed only in brain regions or brain as a whole (Fig. 2A). Specific parts of the brain that reached the significance threshold (FDR 10%) were the limbic system, parietal lobe, brain stem, cerebral cortex, entorhinal cortex, cerebellum, hippocampus, and metencephalon (Fig. 2B). In contrast, for acute back pain, none of the cell types or tissues reached the statistical significance threshold (Fig. 2C), likely due to low overall heritability of acute back pain. Furthermore, the pattern of tissue-specific partitioned heritabilities for acute back pain appeared to be different from those for chronic back pain, with notable absence of significant signals from brain regions (Fig. 2A and 2C).

Next, we used an LD Score regression to estimate genetic correlations between acute and chronic back pain and several selected phenotypes known to be correlated with chronic back pain at the phenotypic level, including body mass index (BMI), insomnia, neuroticism, and depression (Fig. 2D). All of these phenotypes were moderately genetically correlated with chronic back pain, and some of them (BMI, insomnia, and neuroticism) appeared to have significant partitioned heritability in brain tissues, similar to that of chronic back pain (Fig. 2D). In order to look at the similarities/dissimilarities of heritability patterns in brain tissues between chronic back pain and the other phenotypes, we performed a correlation analysis of the heritability estimate coefficients in brain regions (Fig. 2E). A positive correlation would suggest that partitioned heritability is similarly distributed in brain regions between chronic back pain and another phenotype. Interestingly, chronic back pain and acute back pain had strong overall genetic correlation ( $R_g=0.97$ ,  $P<10^{-194}$ ), suggesting that genetic heritability for acute pain, although small, largely overlaps with genetic heritability for chronic back pain. However, when considering only brain tissues, no correlation was found between acute and chronic back pain

( $P=0.53$ , Fig. 2E). Positive correlations of partitioned heritability coefficients in chronic back pain were found for BMI, depression, insomnia, and neuroticism, suggesting that heritability estimates are distributed similarly across brain tissues for these conditions. In contrast, partitioned heritability estimates for brain tissues were not statistically significant in standing height, and did not correlate with heritability estimates for chronic back pain (Fig. 2D-E).

Our findings suggested that chronic back pain is determined by genes predominantly expressed in the brain. To validate our findings, we performed a gene enrichment analysis using data on tissue-specific differential gene expression in mouse assays of inflammatory and neuropathic pain.<sup>35</sup> Four tissues, namely the whole brain, spinal cord, dorsal root ganglia, and whole blood, were available for investigation in two mouse pain assays: Complete Freund Adjuvant (CFA) for inflammatory pain, and Spared Nerve Injury (SNI) for neuropathic pain. In line with our partitioned heritability findings, we observed enrichment of genes upregulated in the nervous system tissues in both mouse pain models (Fig. 3), suggesting that both neuropathic and inflammatory components contribute to human back pain. Inflammatory pain-related genes that were differentially expressed in the brain, spinal cord, and dorsal root ganglia in mice were significantly enriched in chronic, but not acute, back pain GWAS (Fig. 3A). Similarly, neuropathic pain-related genes that were differentially expressed in the mouse brain were significantly enriched in chronic, but not acute, back pain GWAS (FDR 10%, Fig. 3B). Altogether, heritability of chronic back pain is not only largely mediated by SNPs in genes expressed in the brain, but also by those genes that are differentially-expressed in a pain-conditioned brain.

We then hypothesized that epigenetic markers of the 13 genomic loci that reached genome-wide significance (Table 1) in our chronic back pain analyses would be enriched in

brain tissues (Fig. 4). To test this hypothesis, we first mapped the significant SNPs onto corresponding genes and defined 13 gene clusters (Fig. 4A). Then we looked at the intersection of SNPs in linkage disequilibrium (LD) with the lead SNP in each cluster with NIH Epigenetics Roadmap activation markers (H3K4me3 for promoter regions,<sup>36,37</sup> H3K4me1 for enhancers,<sup>36</sup> and H3K36me3 for transcribed regions<sup>38</sup>) for statistically significant enrichment in 396 different tissues<sup>39</sup> (Fig. 4B). Our results again aligned with the partitioned heritability findings: 7 out of 13 loci co-localized with epigenetic markers in multiple brain tissues, as evidenced by statistically significant P-values (Bonferroni-corrected  $P < 0.05$ , Fig. 4B). These results suggest that these regions are transcriptionally active in the CNS. Using a different chromatin epigenome mapping (NIH Epigenetic Roadmap chromatin 15-state model),<sup>40</sup> we found that 9 out of 13 loci had significant enrichment in multiple brain tissues (Fig. 4C). Thus, our epigenetic characterization of the chronic back pain-related variants suggests that expression of the genes in the CNS plays a significant role in the chronic back pain phenotype, followed by genes expressed in the musculoskeletal system (osteoblasts, chondrocytes, and muscles).

## DISCUSSION

Previous studies suggest substantial differences between acute and chronic back pain in terms of predisposing factors.<sup>41,42</sup> Our results from two large cohorts indicate the genetic contribution to chronic back pain is greater than to acute back pain, and much of the heritability of chronic back pain can be traced to genes predominantly expressed in the CNS. At the pathway level, we found the enrichment for genes in the spinal cord ventral commissure morphogenesis pathway in both cohorts. We also mapped the genetic component of chronic but not acute pain states to genes differentially expressed in the brain in mouse pain models. Interestingly, we found this

enrichment for brain specific differentially-expressed genes in both inflammatory and neuropathic mouse pain models, and also in the spinal cord and in dorsal root ganglia of the inflammatory pain model.

Our observation that acute back pain is significantly less heritable than chronic back pain (0.8% for acute versus 4.6% for chronic, narrow-sense SNP heritability) may be due to a greater role of environmental (e.g. tissue injury) factors in acute pain. In turn, transition to chronic back pain, or development of chronic back pain *per se*, appears to require predisposing genetic and/or epigenetic risk factors. It was observed that epigenetic regulation was still occurring in the forebrains of mice that underwent peripheral nerve injuries long after the injury.<sup>43</sup>

While we found a high genetic correlation between chronic and acute back pain, this observation does not contradict our finding that acute back pain is considerably less heritable than chronic back pain. Indeed, genetic correlation is the genetic covariance, i.e., the covariance between sets of SNP contributions to heritability for two traits divided by the square root of  $h_1^2$  and  $h_2^2$ . This genetic correlation can be large if total heritability of one of the traits ( $h_1$ ) due to a set of SNPs is relatively small compared to total heritability of the second trait ( $h_2$ ) and when  $h_1$  constitutes a subset of  $h_2$  in terms of individual SNP contributions. It is inevitable that some individuals with acute back pain in our study will not resolve their pain over time. It is conceivable that the positive genetic correlation may predispose these individuals to chronic back pain. Genetic correlations between chronic back pain and other phenotypes (BMI, insomnia, neuroticism, and depression) were of moderate strength, and are in line with previous studies' results showing genetic similarities between these phenotypes.<sup>26,44,45</sup>

The main finding in our study is that heritability of chronic back pain, but not acute back pain, is largely attributed to genes expressed in brain tissues. Involvement of the CNS in chronic

back pain is an area of active research,<sup>46</sup> supported by a significant comorbidity between chronic back pain and psychological disorders.<sup>47</sup> Indeed, we found genetic overlap of chronic back pain with sleep disorders, neuroticism, and BMI. While this overlap has been previously described in several genetic studies<sup>22,26,48</sup>), here we show that the overlap is localized in genes predominantly expressed in the CNS. Specifically, we found that chronic back pain heritability is mainly localized in the following brain regions: limbic system, parietal lobe, brain stem, cerebral cortex, entorhinal cortex, cerebellum, hippocampus, and metencephalon. Moreover, the distribution of heritability across brain regions was similar for chronic back pain, insomnia, neuroticism, and BMI, suggesting shared genetic and pathophysiological mechanisms between these phenotypes. Our finding of predominant involvement of brain-expressed genes in chronic back pain, but not acute back pain, corroborates previous studies' findings suggesting that spontaneous pain in chronic pain patients involves specific spatiotemporal neuronal mechanisms implicating a salient role for the emotional brain, distinct from mechanisms observed for acute experimental pain.<sup>49</sup> Interestingly, in our study we found only limited evidence of the involvement of musculoskeletal tissues in chronic back pain.

Using the UK Biobank cohort for discovery allowed use of a larger sample size than in previous GWAS studies<sup>26,27</sup> and enabled us to identify 13 loci that exceeded a genome-wide significance threshold for chronic back pain. We have replicated seven loci in the HUNT cohort. Of those, 3 loci are novel (Chr5: LOC105374704 • CDH6, Chr5: NUDT12 • RAB9BP1, Chr14: SLC25A21-AS1 • MIPOL1 • FOXA1). Another locus (Chr18: LINC01630 • DCC • LINC01919) has never previously been replicated in an independent cohort.

Our results suggest that chronic pain is much more heritable than acute pain. Furthermore, despite of the strong genetic correlation between acute and chronic pain, the most

significant pathways contributing to these pain states seem to be different. Pathway analyses of chronic back pain GWAS revealed enrichment for genes involved in the spinal cord ventral commissure morphogenesis pathway in both the UK Biobank and HUNT cohorts. The ventral white commissure is comprised of A delta and C fibers, which are known to be involved in pain signal transduction. This suggests that molecular mechanisms of nerve fiber growth might be involved in the pathophysiology of chronic back pain.<sup>50</sup> Interestingly, the most significant pathway in acute back pain, GO:00042483, contains two genes linked to connective tissue disorders (RSPO2) and bone remodeling (TNFRSF11B).<sup>31,32 33</sup> It is possible, therefore, that different gene subsets drive the bulk of the onset and persistence of back pain, with the majority of the pathways contributing to acute back pain not necessarily contributing to pain becoming chronic, in addition to a strong environmental component in acute back pain.

A limitation of our study is in the extent to which the phenotypes of interest, acute and chronic back pain, are ascertained in the UK Biobank cohort. Pain intensity, frequency of episodes, and pain medication use were not available for analyses. Additionally, phenotype definitions varied slightly between the UK Biobank and the HUNT cohorts (back pain in the UK Biobank and low back pain in the HUNT). However, we concluded that these differences did not affect the validity of our findings given the high replication success rate of our findings between the two cohorts.

In conclusion, our analyses show that chronic back pain is substantially more heritable than acute back pain, and this heritability is mostly attributed to genes expressed in the brain. Furthermore, our results suggest that different genetic pathways may be responsible for acute and chronic pain manifestation. Molecular pathophysiology of acute back pain is largely contributed to by connective and bone tissue remodeling and inflammatory pathways while chronic back

pain is contributed to by neuronal development processes. These results provide substantial insight into physiological systems and pathways responsible for acute and chronic back pain, and should be expanded upon to further to develop targeted treatments.

## **METHODS**

### **The UK Biobank cohort**

The UK Biobank cohort was used as a discovery cohort. This cohort is a high-powered prospective study of 500,000 people recruited in the United Kingdom.<sup>51,52</sup> It is mainly comprised of individuals of Caucasian ancestry whose ages range between 37 and 73 years, with a female to male participant ratio of about 1.2:1. Chronic cases were defined as those answering “yes” to the question “Have you had back pains for more than 3 months?” (field 3571). This question was asked to participants if they answered “yes” for the “Back Pain” option at this question: “In the last month have you experienced any of the following that interfered with your usual activities?” (field 6159). Participants were discarded based on the following criteria: answering “no”, “do not know” or “prefer not to answer” (field 3571), not “White British”, failed genotyping quality control or sex mismatch, and voluntary withdrawal from the study. Acute back pain cases were identified as those answering “no” at field 3571. All other participants (i.e. those who did not meet the criteria for acute or chronic back pain) were qualified as controls. Therefore, control individuals might have experienced acute or chronic pain at other bodily sites than in the back. Genotyping, quality control, and genomic imputation in the UK Biobank cohort were previously described.<sup>53</sup>

### **The HUNT cohort**

The Nord-Trøndelag Health Study (HUNT) cohort was used for replication of results. The HUNT study has been conducted in Nord-Trøndelag County, Norway, in three consecutive waves<sup>54</sup>: HUNT in 1984–1986, HUNT2 in 1995–1997 and HUNT3 in 2006–2008. This work combines baseline data from the second survey, HUNT2,<sup>55</sup> and the third survey, HUNT3.<sup>54</sup> All residents of Norway aged 20 years and older were invited to take part in the HUNT2 survey. They were requested to complete a questionnaire on health status and invited to a clinical consultation that included measuring height and weight. In the HUNT3 survey 11 years later, similar information was collected using a questionnaire and a clinical examination. Each participant in the HUNT2 and HUNT3 surveys signed a written informed consent regarding the collection and use of data for research purposes.

One question in the HUNT2 and HUNT3 questionnaires was expressed: ‘During the last year, have you suffered from pain and/or stiffness in your muscles and joints that has lasted for at least 3 consecutive months?’ Each participant answering yes was given the following question: ‘Where did you have these complaints?’ Several body regions were listed. Individuals answering yes to the first question and including the lower back as a relevant region were regarded as having chronic lower back pain.<sup>56</sup> Acute low back pain cases (n=4,379) were defined as those who had lumbar pain in the last month (at HUNT2), but not had musculoskeletal pain for more than 3 of the past 12 months (at HUNT2). Controls (n=11,309) were defined as those participants without musculoskeletal pain in the last month or musculoskeletal pain for more than 3 of the past 12 months (at HUNT2) and if they did not have musculoskeletal pain for more than 3 of the past 12 months (at HUNT3). Chronic low back pain cases (n=19,760) were defined as those who had lumbar pain for more than 3 of the past 12 months. Controls (n=28,674) had no musculoskeletal pain for more than 3 of the past 12 months at either HUNT2 or HUNT3.

Genotyping of study subjects was performed in three batches using the Illumina HumanCoreExome (Illumina Inc, CA, USA).

### **Statistical analyses**

Genome-wide association tests in the UK Biobank cohort were performed using a linear mixed model as implemented in BOLT-LMM software, version 2.3.<sup>29,30</sup> Covariates were sex, age, age<sup>2</sup>, genotyping platform, first 40 genetic principal components, and recruitment centers. Kinship was considered by BOLT-LMM using genotyped position data. Retained SNPs had minor allele frequencies of at least one-in-a-thousand, departed from Hardy-Weinberg equilibrium with *P*-values greater or equal to 10<sup>-12</sup>, and were part of the Haplotype Reference Consortium (HRC) panel.<sup>57</sup> Association analyses in the HUNT cohort were performed using a mixed logistic regression (SAIGE), adjusted for sex, genotyping batch, and 5 principal components.

Epigenetic data for functional characterization of GWAS significant SNPs was taken from the NIH Roadmap Epigenomics Consortium.<sup>39,40</sup> At each significant locus, the list of SNPs in high LD ( $r^2 \geq 0.5$ ) with the locus' lead SNP was retrieved from LDlink<sup>58</sup> using the Great Britain population as the reference. We used the 'intersect' option of the bedtools<sup>59</sup> to retrieve high-LD SNPs that overlapped with epigenetic features. Two analyses of epigenetic markers were performed: the first one in relation to the overlap of high-LD SNPs with known epigenetic activation marker peaks, and the second one in relation to the overlap with NIH Roadmap's 15-states chromatin model built from peaks of activation/repression markers, focusing on transcription start sites (state 1), transcription enhancers (states 6 or 7) indicating active transcription, and bivalent or poised transcription start site (state 10) indicating absence of active

transcription. For analyses of activation peaks, retained broad peaks had experimental ChIP-Seq evidence P-values  $\leq (0.05 / (13 \times 19 \times 3))$ , correcting for 13 genome-wide significant loci in 19 pain-relevant tissues for 3 epigenetic activation markers (H3K4me3, H3K4me1, and H3K36me3). For chromatin states, we only considered the presence of an overlap with one of these states (since no P-values were available), prioritized as follows: state 1; states 6 and 7; state 10. All data files were downloaded from the following URLs:

<https://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/gappedPeak/>,

<https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final/>.

Partitioned heritability analyses were performed to determine heritability enrichment in annotated SNP sets using an LD Score Regression (LDSC).<sup>60</sup> Functional genomic annotations (tissue- and cell-specific gene expression) were retrieved from an online source database comprised of 152 different human tissues.<sup>34</sup>

Additional custom annotations were built from genes overexpressed in two mouse pain assays.<sup>35</sup> The mouse pain assays were Complete Freund's Adjuvant, a model for inflammatory pain, and Spared Nerve Injury, a model for neuropathic pain. Quantitation of gene expression from the RNA-Seq data (GEO accession GSE111216) was obtained using STAR<sup>61</sup> on Genome Reference Consortium's GRCm38 genome, followed by featureCount<sup>62</sup> to extract reads per genes. Over-expressed genes in the two pain assays compared to naïve were identified via DESeq2.<sup>63</sup> A total of three replicate samples per condition (naïve, CFA, SNI) were available. Mouse gene names were translated to human gene names using the BioMart tool.<sup>64</sup> The top 10% (n=1,663) genes with the greatest positive test statistic (more expression in the pain assay) were

retained to build LDSC's genetic annotation track, as recommended.<sup>60</sup> The reference or null model annotation track was built from all annotated genes (n=16,629).

For pathway analyses, the summary GWAS data were first analyzed using MAGMA,<sup>65,66</sup> which aggregated GWAS SNP P-values into gene-level ones, while considering linkage-disequilibrium between the SNPs. MAGMA was also used to deduce pathway-level P-values from gene-level ones. Pathways were sourced from Gene Ontology's biological processes,<sup>67,68</sup> obtained from the Bader lab at URL: [http://download.baderlab.org/EM\\_Genesets/December\\_01\\_2019/](http://download.baderlab.org/EM_Genesets/December_01_2019/). Pathways in the discovery cohort (the UK Biobank) at the FDR 20% level were ascertained for replication in the replication cohort (HUNT) with P-value of replication smaller or equal to 0.05.

**Author contributions:** LD and AB conceived the paper; AB, MP, and LD designed the analytical plan and experiments. MP, SK, and AEM performed the bioinformatics analyses. AB and MP interpreted the results and wrote the manuscript. AEM, MUL, IH, KH, JAZ, BSW provided the data from the HUNT study. SK, DZ, AEM, MUL, IH, KH, JAZ, BSW, and LD critically reviewed the manuscript. All the authors read and approved the final manuscript.

**Acknowledgments:** We would like to thank all participants enrolled in the studies herein. The analysis in the UK Biobank was performed under application 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).

**Group Members “HUNT All-In Pain”:** Anne Heidi Skogholt<sup>1,2</sup>, Ben Brumpton<sup>1</sup>, Cristen Willer<sup>3</sup>, Egil Andreas Fors<sup>4</sup>, Ingunn Mundal<sup>5</sup>, Jonas Bille Nielsen<sup>1,3,6</sup>, Kjersti Storheim<sup>7,8</sup>, Knut Hagen<sup>9</sup>, Kristian Bernhard Nilsen<sup>9,10</sup>, Lars Fritsche<sup>11</sup>, Laurent F. Thomas<sup>1,2,12,13</sup>, Linda M Pedersen<sup>14</sup>, Maiken E Gabrielsen<sup>1</sup>, Marianne Bakke Johnsen<sup>1,7,15</sup>, Oddgeir Holmen<sup>16</sup>, Sigrid Børte<sup>1,7,15</sup>, Synne Øien Stensland<sup>7,17</sup>, Wei Zhou<sup>18,19</sup>. **Affiliations:** <sup>1</sup> 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. <sup>2</sup> Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway. <sup>3</sup> Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, 48109, MI, USA. <sup>4</sup> Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. <sup>5</sup> Department of Health Science, Molde University College, Molde, Norway. <sup>6</sup> Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. <sup>7</sup> Research and Communication Unit for Musculoskeletal Health (FORMI), Department of Research, Innovation and Education, Division of Clinical

Neuroscience, Oslo University Hospital, Oslo, Norway.<sup>8</sup> Faculty of Health Sciences, Department of physiotherapy, Oslo Metropolitan University, Oslo, Norway.<sup>9</sup> Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

<sup>10</sup> Department of Neurology, Oslo University Hospital, Oslo, Norway. <sup>11</sup> Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, 48109, MI, USA. <sup>12</sup> BioCore - Bioinformatics Core Facility, Norwegian University of Science and Technology, Trondheim, Norway. <sup>13</sup> Clinic of Laboratory Medicine, St.Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. <sup>14</sup> Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway. <sup>15</sup> Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway. <sup>16</sup> HUNT Research Center, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. <sup>17</sup> NKVTS, Norwegian Centre for Violence and Traumatic Stress Studies. <sup>18</sup> Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. <sup>19</sup> Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA.

**Competing interests:** The authors declare no competing financial interests.

**Funding:** Funding for this work was kindly provided by the Canadian Excellence Research Chairs (CERC) Program ([www.cerc.gc.ca](http://www.cerc.gc.ca)) grant CERC09 (to LD). DVZ was supported in part by the Intramural Research Program of the NIH.



**Table 1.** SNP-level discovery and replication results of chronic back pain GWAS. Replication of the top hit within each locus was performed using SNP sets in high LD ( $r^2 > 0.8$ ). SNPs with lowest replication P-values for each locus in the discovery and replication cohort are shown.

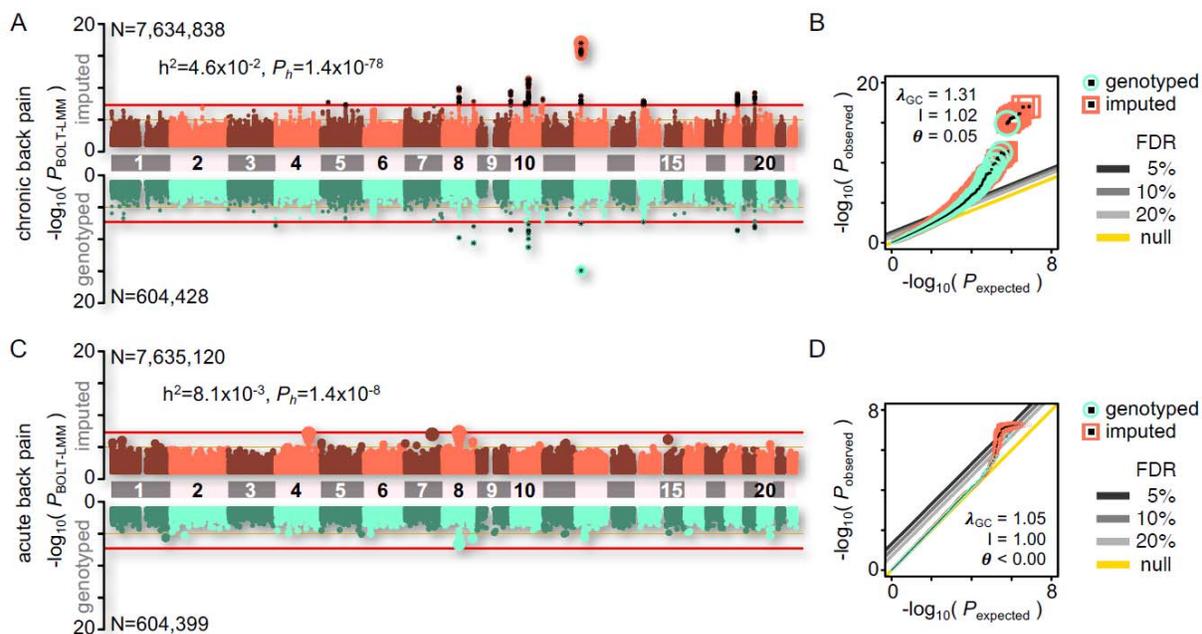
	Chr	Genes Cluster	Discovery, top hit			Replication				
			SNP	BP	P	SNP	BP	P	r2	Direction
1	4	LOC105374344•TMED11P•SPON2	rs6826705	1122634	1.30E-08	rs4368552	1128054	0.067	0.97	↑↑
2	5	LOC105374704•CDH6	rs2066928	30843787	1.80E-08	rs2066928	30843787	0.034	1	↑↑
3	5	NUDT12•RAB9BP1	rs325485	103995368	3.90E-08	rs325528	104048590	0.0065	0.88	↑↑
4	8	C8orf34-AS1•C8orf34•LINC01592	rs1865442	69574165	1.10E-10	rs72666774	69587018	0.0042	0.96	↑↑
5	8	CCDC26•GSDMC	rs12056383	130711839	1.00E-08	rs77753889	130711599	0.631	0.85	↑↓
6	9	NRARP•EXD3•NOXA1	rs28458909	140257189	2.90E-09	--	--	--	--	--
7	10	LINC00844•CCEPR	rs12415581	60899641	2.10E-08	rs7084967	60945358	0.356	0.80	↑↑
8	10	PSAP•CHST3•SPOCK2•ASCC1	rs1668172	73823019	4.10E-12	rs1245561	73837350	0.0046	0.98	↑↑
9	10	BNIP3•JAKMIP3•DPYSL4	rs10870267	133968063	5.00E-09	rs34588848	134005567	0.480	0.83	↑↑
10	12	LOC101928441•SOX5•SOX5-AS1	rs12310519	23975219	9.50E-18	rs56290807	23972014	0.035	0.97	↑↑
11	14	SLC25A21-AS1•MIPOL1•FOXA1	rs8018823	37718177	9.70E-09	rs8018493	37644792	0.012	0.82	↑↑
12	18	LINC01630•DCC•LINC01919	rs10502966	50748499	8.40E-10	rs17410557	50776391	2.33E-05	0.83	↑↑
13	19	BCAM•NECTIN2•TOMM40•APOE•APOC1•APOC1P1	rs12972970	45387596	2.50E-09	rs34342646	45388130	0.363	1	↑↑

**Table 2.** Pathway analysis in the discovery (UK Biobank) and replication (HUNT) cohorts. Pathways that pass the significance threshold FDR 20% in the discovery cohort are shown.

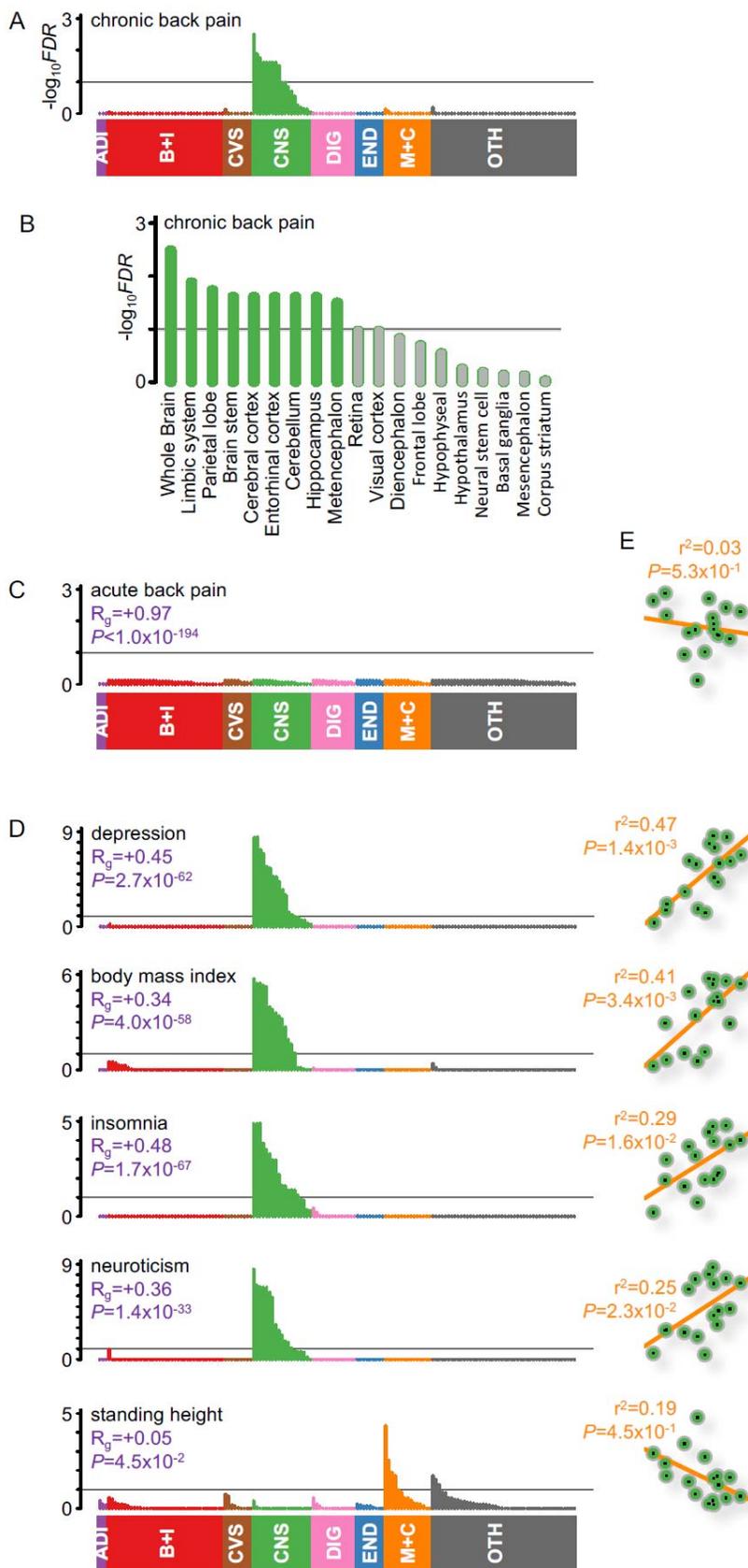
GO term	Description	Leading edge genes	Discovery FDR	Replicated at $P < 0.05$ ?
<b>Chronic back pain</b>				
GO:0021965	spinal cord ventral commissure morphogenesis*	DCC,ADARB1,GLI2	9.06E-05	Yes
GO:0099560	synaptic membrane adhesion	NRXN1,EFNA5,LRFN5,LRRRC4C,IGSF9B,NLGN1,CDH9,NTNG1,PCDH17,LRRRC4B	1.85E-02	No
GO:0021955	central nervous system neuron axonogenesis	DCC,PTK2,TSKU,ADARB1,HSP90AB1,ARHGAP35,NDEL1,DCLK1,SLIT2,ZEB2	1.85E-02	No
GO:0051347	positive regulation of transferase activity*	FGFR1,F2,AMBRA1,ATG13,PTK2,EFNA5,DGKZ,FYN,STOX1,FGF18	1.30E-01	Yes
GO:1990074	polyuridylation-dependent mRNA catabolic process	ZCCHC11,DIS3L2,ZCCHC6	1.30E-01	No
GO:0060684	epithelial-mesenchymal cell signaling	FOXA1,BMP4,PDGFA,SHH,WNT6,SMO	1.30E-01	No
GO:0033674	positive regulation of kinase activity*	FGFR1,F2,AMBRA1,ATG13,PTK2,EFNA5,DGKZ,FYN,STOX1,FGF18	1.34E-01	Yes
GO:0021952	central nervous system projection neuron axonogenesis	DCC,TSKU,ADARB1,DCLK1,SLIT2,ZEB2,NR2E1,CDH11,NFIB,GLI2	1.45E-01	No
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	SDK1,ADGRL3,NRXN1,EFNA5,ROBO2,LRFN5,PCDHB12,CDHR4,REG3A,PLXNB2	1.51E-01	No
GO:0021670	lateral ventricle development	TSKU,NUMB,ATP1B2,MYH10,DPCD,KDM2B,CDK6,PAX5,DNAH5,NUMBL	1.51E-01	No
GO:0060255	regulation of macromolecule metabolic process	SPOCK2,FNIP2,MAML3,BBX,FGFR1,EFNB2,FAM172A,F2,PHF21A,SP4	1.51E-01	No
GO:0051963	regulation of synapse assembly	ADGRL3,NRXN1,PTK2,EFNA5,ROBO2,LRFN5,NTRK1,NLGN2,MUSK,AMIGO3	1.51E-01	No
GO:0031325	positive regulation of cellular metabolic process	FNIP2,MAML3,FGFR1,F2,SP4,AMBRA1,ATG13,APOC1,PTK2,EFNA5	1.51E-01	No
GO:0019222	regulation of metabolic process	SPOCK2,FNIP2,MAML3,BBX,FGFR1,EFNB2,FAM172A,F2,PHF21A,SP4	1.51E-01	No
GO:1902841	regulation of netrin-activated signaling pathway	DCC,SIAH2,SIAH1,NTN1	1.51E-01	No
GO:1902842	negative regulation of netrin-activated signaling pathway	DCC,SIAH2,SIAH1,NTN1	1.51E-01	No
GO:0098609	cell-cell adhesion	SDK1,ADGRL3,NRXN1,EFNA5,CYFIP2,THBS4,ROBO2,LRFN5,NLGN2,PCDHB12	1.53E-01	No
GO:0051128	regulation of cellular component organization	DCC,SDK1,FNIP2,ADGRL3,FGFR1,CKAP5,NRXN1,EFNB2,F2,ATG13	1.53E-01	No
GO:0097116	gephyrin clustering involved in postsynaptic density assembly	NRXN1,NLGN2,NRXN2	1.53E-01	No
GO:0030702	chromatin silencing at centromere	HIRA,ZNFX1	1.53E-01	No
GO:0010604	positive regulation of macromolecule metabolic process	FNIP2,MAML3,FGFR1,F2,SP4,ATG13,PTK2,EFNA5,SPON1,FYN	1.82E-01	No
<b>Acute back pain</b>				
GO:0042483	negative regulation of odontogenesis	RSPO2,TNFRSF11B,ASPN	3.05E-02	No

GO:1904381	Golgi apparatus mannose trimming	MAN1A1,MAN1A2,MAN1C1	1.24E-01	No
GO:0072033	renal vesicle formation	WNT4,FMN1,KIF26B,SMO,CTNNB1,SALL1	1.30E-01	No
GO:0042481	regulation of odontogenesis	RSPO2,TNFRSF11B,WNT10A,DMP1,ASPN,ENAM,C4orf26,MMP20,SP6,PAX9	1.30E-01	No
GO:0098914	membrane repolarization during atrial cardiac muscle cell action potential	KCNJ5,KCNJ3,KCNN2,KCNQ1,KCNA5	1.30E-01	No
GO:0099624	atrial cardiac muscle cell membrane repolarization	KCNJ5,KCNJ3,KCNN2,KCNQ1,KCNA5	1.30E-01	No
GO:0015838	amino-acid betaine transport	PDZK1,SLC38A2,SLC25A29,SLC22A4,SLC22A5,SLC22A16	1.30E-01	No
GO:0042489	negative regulation of odontogenesis of dentin-containing tooth	RSPO2,TNFRSF11B	1.30E-01	No
GO:0051409	response to nitrosative stress	ATG5,STOX1,GCLC,GCLM,DUSP6,ADH5,DDIT3,ATM	1.51E-01	No
GO:0045630	positive regulation of T-helper 2 cell differentiation	CD86,NLRP3,IL4R,PRKCZ,IL18,TNFSF4,RARA	1.65E-01	No

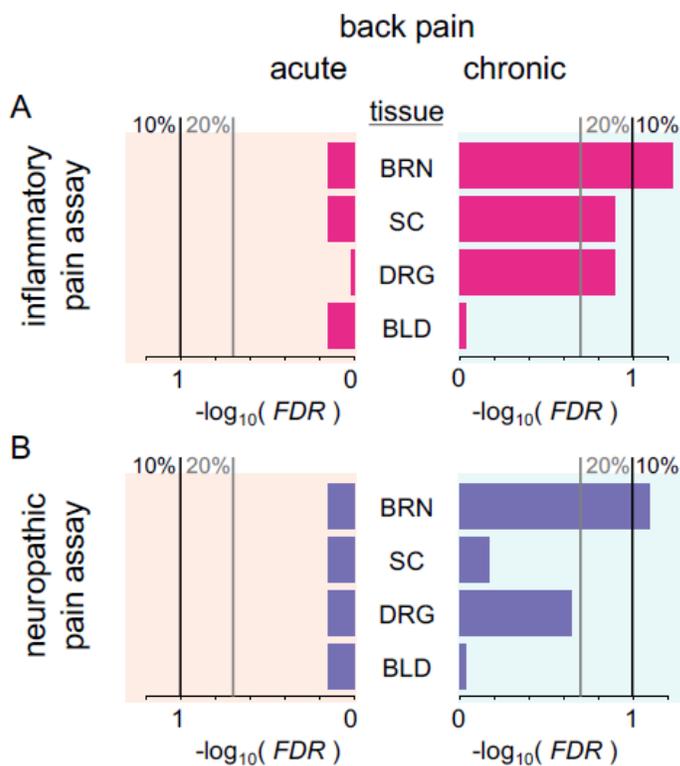
## FIGURES AND FIGURE LEGENDS



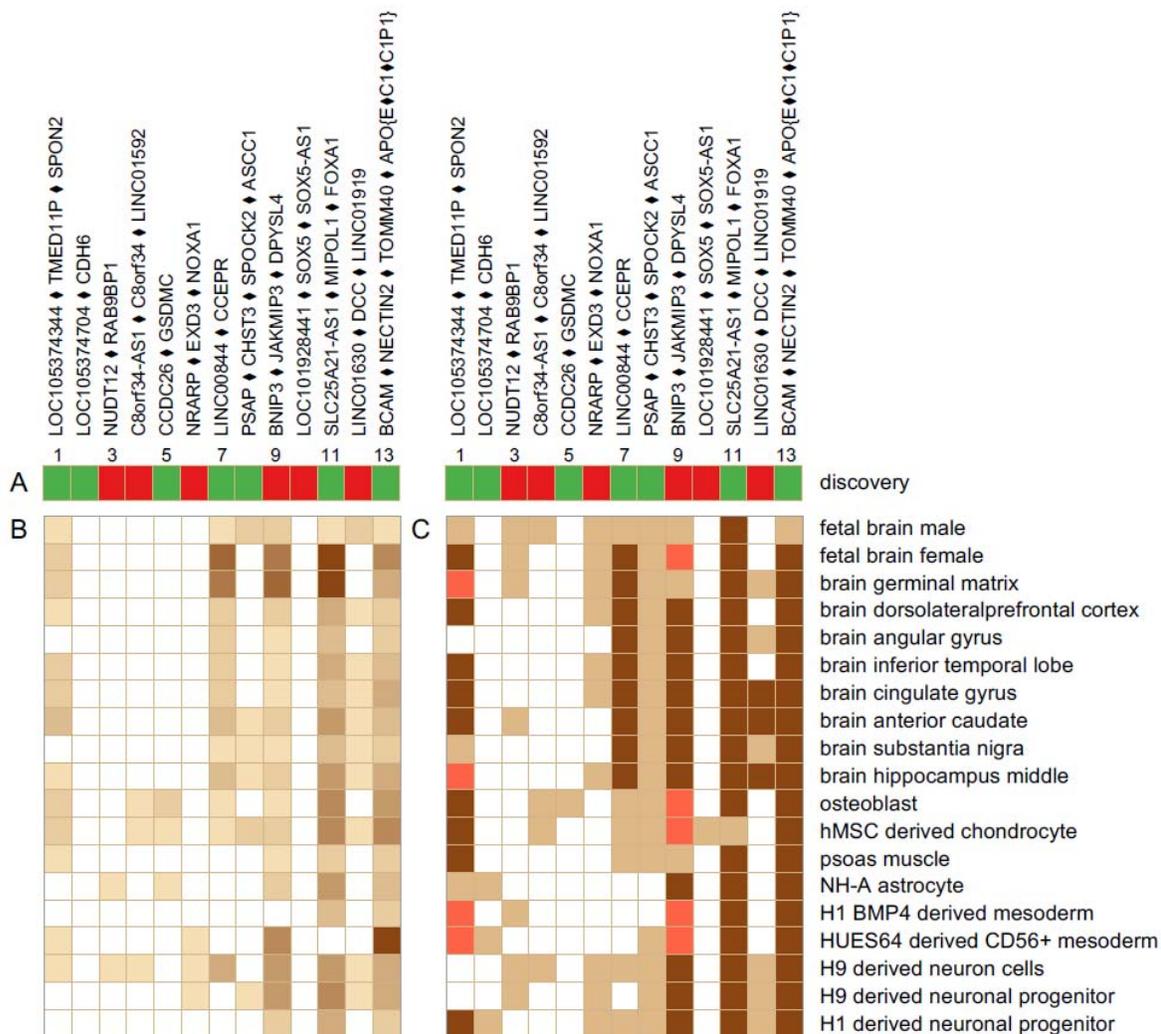
**Figure 1.** Genome-wide association analysis of chronic back pain (A, B) and acute back pain (C, D) in 375,158 participants from the UK Biobank cohort. (A, C) Miami plots show  $-\log_{10}$  P-values for imputed (top, orange) and genotyped (bottom, green) SNPs (N, number of analyzed subjects). Horizontal red lines indicate genome-wide statistical significance at  $5 \times 10^{-8}$ . Genetic heritability estimate ( $h^2$ ) shown with corresponding P-value ( $P_h$ ). (B, D) QQ plots show observed  $-\log_{10}$  P-values ( $P_{\text{observed}}$ ) as a function of expected  $-\log_{10}$  P-values ( $P_{\text{expected}}$ ), for both genotyped (green) and imputed (orange) SNPs. Genomic control parameter ( $\lambda_{\text{GC}}$ ), LD Score regression intercept (I) and their ratios ( $\theta$ ) are shown.



**Figure 2.** Tissue-specific heritability enrichment for chronic back pain, acute back pain, and other related phenotypes. Genomic tissue-specific annotations for 152 tissues were grouped into eight categories: adipose, ADI (n=3, purple); blood + immune, B+I (n=37, red), cardiovascular, CVS (n=9, brown); central nervous system, CNS (n=19, green); digestive, DIG (n=14, pink); endocrine, END (n=9, blue); musculoskeletal + connective, M+C (n=15, orange); and other, OTH (n=46, grey). Vertical bars in each plot denote  $-\log_{10}(\text{FDR})$  values for enrichment of each tissue. FDR 10% threshold is denoted by a horizontal grey line. **(A, B)** Tissue-specific partitioned heritability for chronic back pain. **(C)** Tissue-specific partitioned heritability for acute back pain, genetic correlation ( $R_g$ ) with chronic back pain, and genetic correlation P-value. **(D)** Tissue-specific partitioned heritability for depression, body mass index, insomnia, neuroticism and standing height, genetic correlation ( $R_g$ ) with chronic back pain and genetic correlation P-value. **(E)** Brain regions partitioned heritability enrichment correlation between chronic back pain (horizontal axis, not shown) and acute back pain, depression, body mass index, insomnia, neuroticism, and standing height (vertical axis, not shown). Squared correlation coefficient ( $r^2$ ) and associated P-value ( $P$ ) are shown in orange.



**Figure 3.** Tissue-specific heritability enrichment for back pain GWAS results using transcriptomic annotations from inflammatory (A) and neuropathic (B) mouse pain assays. Annotations for genes upregulated in whole brain (BRN), spinal cord (SC), dorsal root ganglia (DRG), and whole blood (BLD) were used to determine heritability enrichment in chronic and acute back pain GWAS. Shown are  $-\log_{10}$  FDR-corrected P-values for enrichment of heritability. Vertical black lines indicate FDR 10%; vertical grey lines indicate FDR 20%.



**Figure 4.** Epigenetic characterization of the GWAS significant 13 chronic back pain loci. **(A)** Gene clusters. Lead SNP effect of minor allele in the discovery cohort: protective (green) or risk (red). **(B)** Intersection of SNPs in LD with lead SNP ( $r^2 \geq 0.5$ ) with NIH Epigenetics Roadmap activation markers in selected tissues. Markers are: H3K4me3, H3K4me1, H3K36me3. Darker hues for increased signal. **(C)** Intersection of SNPs in LD with lead SNP ( $r^2 \geq 0.5$ ) with NIH Epigenetics Roadmap 15-states chromatin model in selected tissues. States are: transcription start site (state 1; dark brown) and enhancer (states 6 or 7; light brown) indicating active transcription, and bivalent or poised transcription start site (state 10; orange) indicating absence of active

transcription. Statistical significance for ChIP-Seq signal established at  $P \leq 0.05/(3 \text{ markers} \times 13 \text{ loci} \times 19 \text{ tissues})$ . Cells are colored white otherwise.

## REFERENCES

- 1 Dagenais, S., Caro, J. & Haldeman, S. A systematic review of low back pain cost of illness studies in the United States and internationally. *Spine J* **8**, 8-20, doi:10.1016/j.spinee.2007.10.005 (2008).
- 2 Lim, K.-L., Jacobs, P. & Klarenbach, S. A population-based analysis of healthcare utilization of persons with back disorders: results from the Canadian Community Health Survey 2000–2001. *Spine* **31**, 212-218 (2006).
- 3 Von Korff, M. *et al.* Chronic spinal pain and physical–mental comorbidity in the United States: results from the national comorbidity survey replication. *Pain* **113**, 331-339 (2005).
- 4 Coyte, P. C., Asche, C. V., Croxford, R. & Chan, B. The economic cost of musculoskeletal disorders in Canada. *Arthritis & Rheumatism* **11**, 315-325 (1998).
- 5 Deyo, R. A. *et al.* Report of the NIH Task Force on research standards for chronic low back pain. *J Pain* **15**, 569-585, doi:10.1016/j.jpain.2014.03.005 (2014).
- 6 Cassidy, J. D., Côté, P., Carroll, L. J. & Kristman, V. Incidence and course of low back pain episodes in the general population. *Spine* **30**, 2817-2823 (2005).
- 7 Hancock, M. J. *et al.* Assessment of diclofenac or spinal manipulative therapy, or both, in addition to recommended first-line treatment for acute low back pain: a randomised controlled trial. *The Lancet* **370**, 1638-1643 (2007).
- 8 Coyte, P. C., Asche, C. V., Croxford, R. & Chan, B. The economic cost of musculoskeletal disorders in Canada. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* **11**, 315-325 (1998).

- 9 Krismer, M. & Van Tulder, M. Strategies for prevention and management of musculoskeletal conditions. *Best Pract Res Clin Rheumatol* **21**, 77-91 (2007).
- 10 Correa, J. B., Costa, L. O., de Oliveira, N. T., Sluka, K. A. & Liebano, R. E. Central sensitization and changes in conditioned pain modulation in people with chronic nonspecific low back pain: a case-control study. *Exp Brain Res* **233**, 2391-2399, doi:10.1007/s00221-015-4309-6 (2015).
- 11 Kreiner, D. S. *et al.* An evidence-based clinical guideline for the diagnosis and treatment of lumbar disc herniation with radiculopathy. *Spine J* **14**, 180-191, doi:10.1016/j.spinee.2013.08.003 (2014).
- 12 van Kleef, M. *et al.* Pain originating from the lumbar facet joints. *Pain Pract* **10**, 459-469, doi:10.1111/j.1533-2500.2010.00393.x (2010).
- 13 Patel, V. Sacroiliac joint injuries. *Orthopedics* **35**, 398, doi:10.3928/01477447-20120426-05 (2012).
- 14 Manchikanti, L. *et al.* An update of comprehensive evidence-based guidelines for interventional techniques in chronic spinal pain. Part II: guidance and recommendations. *Pain Physician* **16**, S49-283 (2013).
- 15 Dreyfuss, P., Dreyer, S. J., Cole, A. & Mayo, K. Sacroiliac joint pain. *J Am Acad Orthop Surg* **12**, 255-265, doi:10.5435/00124635-200407000-00006 (2004).
- 16 Ciricillo, S. F. & Weinstein, P. R. Lumbar spinal stenosis. *West J Med* **158**, 171-177 (1993).
- 17 Zheng, C. J. & Chen, J. Disc degeneration implies low back pain. *Theor Biol Med Model* **12**, 24, doi:10.1186/s12976-015-0020-3 (2015).

- 18 Battie, M. C., Videman, T., Levalahti, E., Gill, K. & Kaprio, J. Heritability of low back pain and the role of disc degeneration. *Pain* **131**, 272-280, doi:10.1016/j.pain.2007.01.010 (2007).
- 19 Junqueira, D. R. *et al.* Heritability and lifestyle factors in chronic low back pain: results of the Australian twin low back pain study (The AUTBACK study). *Eur J Pain* **18**, 1410-1418, doi:10.1002/ejp.506 (2014).
- 20 MacGregor, A. J., Andrew, T., Sambrook, P. N. & Spector, T. D. Structural, psychological, and genetic influences on low back and neck pain: a study of adult female twins. *Arthritis Rheum* **51**, 160-167, doi:10.1002/art.20236 (2004).
- 21 Nyman, T., Mulder, M., Iliadou, A., Svartengren, M. & Wiktorin, C. High heritability for concurrent low back and neck-shoulder pain: a study of twins. *Spine (Phila Pa 1976)* **36**, E1469-1476, doi:10.1097/BRS.0b013e3181e2c878 (2011).
- 22 Pinheiro, M. B. *et al.* Genetics and the environment affect the relationship between depression and low back pain: a co-twin control study of Spanish twins. *Pain* **156**, 496-503, doi:10.1097/01.j.pain.0000460330.56256.25 (2015).
- 23 Zadro, J. R. *et al.* Does educational attainment increase the risk of low back pain when genetics are considered? A population-based study of Spanish twins. *Spine J* **17**, 518-530, doi:10.1016/j.spinee.2016.10.021 (2017).
- 24 Dario, A. B. *et al.* The relationship between obesity, low back pain, and lumbar disc degeneration when genetics and the environment are considered: a systematic review of twin studies. *Spine J* **15**, 1106-1117, doi:10.1016/j.spinee.2015.02.001 (2015).
- 25 Malkin, I. *et al.* Low back and common widespread pain share common genetic determinants. *Ann Hum Genet* **78**, 357-366, doi:10.1111/ahg.12074 (2014).

- 26 Freidin, M. B. *et al.* Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individuals. *Pain* **160**, 1361-1373, doi:10.1097/j.pain.0000000000001514 (2019).
- 27 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, doi:10.1371/journal.pgen.1007601 (2018).
- 28 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295, doi:10.1038/ng.3211 (2015).
- 29 Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat Genet* **50**, 906-908, doi:10.1038/s41588-018-0144-6 (2018).
- 30 Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-290, doi:10.1038/ng.3190 (2015).
- 31 Nakajima, M., Kou, I., Ohashi, H., Genetic Study Group of the Investigation Committee on the Ossification of Spinal, L. & Ikegawa, S. Identification and Functional Characterization of RSPO2 as a Susceptibility Gene for Ossification of the Posterior Longitudinal Ligament of the Spine. *Am J Hum Genet* **99**, 202-207, doi:10.1016/j.ajhg.2016.05.018 (2016).
- 32 Ten Dam, E. J., van Beuge, M. M., Bank, R. A. & Werker, P. M. Further evidence of the involvement of the Wnt signaling pathway in Dupuytren's disease. *J Cell Commun Signal* **10**, 33-40, doi:10.1007/s12079-015-0312-8 (2016).

- 33 Ren, X. *et al.* Osteoprotegerin reduces osteoclast resorption activity without affecting osteogenesis on nanoparticulate mineralized collagen scaffolds. *Sci Adv* **5**, eaaw4991, doi:10.1126/sciadv.aaw4991 (2019).
- 34 Fehrmann, R. S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47**, 115-125, doi:10.1038/ng.3173 (2015).
- 35 Parisien, M. *et al.* Genetic pathway analysis reveals a major role for extracellular matrix organization in inflammatory and neuropathic pain. *Pain* **160**, 932-944, doi:10.1097/j.pain.0000000000001471 (2019).
- 36 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318, doi:10.1038/ng1966 (2007).
- 37 Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169-181, doi:10.1016/j.cell.2005.01.001 (2005).
- 38 Bonasio, R., Tu, S. & Reinberg, D. Molecular signals of epigenetic states. *Science* **330**, 612-616, doi:10.1126/science.1191078 (2010).
- 39 Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* **28**, 1045-1048, doi:10.1038/nbt1010-1045 (2010).
- 40 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
- 41 Gatchel, R. J., Polatin, P. B. & Mayer, T. G. The dominant role of psychosocial risk factors in the development of chronic low back pain disability. *Spine (Phila Pa 1976)* **20**, 2702-2709, doi:10.1097/00007632-199512150-00011 (1995).

- 42 Jayson, M. Chronic back pain is not the same as acute back pain lasting longer. *BMJ* **314**, 1639-1640 (1997).
- 43 Tajerian, M. *et al.* Peripheral nerve injury is associated with chronic, reversible changes in global DNA methylation in the mouse prefrontal cortex. *PLoS One* **8**, e55259, doi:10.1371/journal.pone.0055259 (2013).
- 44 Ropponen, A. *et al.* Personality traits and life dissatisfaction as risk factors for disability pension due to low back diagnoses: A 30-year longitudinal cohort study of Finnish twins. *Journal of psychosomatic research* **73**, 289-294 (2012).
- 45 Suri, P. *et al.* Modifiable risk factors for chronic back pain: insights using the co-twin control design. *The Spine Journal* **17**, 4-14 (2017).
- 46 Schoupe, S., Van Oosterwijck, S., Danneels, L., Van Damme, S. & Van Oosterwijck, J. Are Functional Brain Alterations Present in Low Back Pain? A Systematic Review of EEG Studies. *J Pain*, doi:10.1016/j.jpain.2019.06.010 (2019).
- 47 Vachon-Preseau, E. *et al.* Identification of traits and functional connectivity-based neurotraits of chronic pain. *PLoS Biol* **17**, e3000349, doi:10.1371/journal.pbio.3000349 (2019).
- 48 Pinheiro, M. B. *et al.* Genetic and environmental contributions to sleep quality and low back pain: a population-based twin study. *Psychosomatic medicine* **80**, 263-270 (2018).
- 49 Baliki, M. N. *et al.* Chronic pain and the emotional brain: specific brain activity associated with spontaneous fluctuations of intensity of chronic back pain. *J Neurosci* **26**, 12165-12173, doi:10.1523/JNEUROSCI.3576-06.2006 (2006).
- 50 Cross, S. A. in *Mayo Clinic Proceedings*. 375-383 (Elsevier).

- 51 Allen, N. E., Sudlow, C., Peakman, T., Collins, R. & Biobank, U. K. UK biobank data: come and get it. *Sci Transl Med* **6**, 224ed224, doi:10.1126/scitranslmed.3008601 (2014).
- 52 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, doi:10.1371/journal.pmed.1001779 (2015).
- 53 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).
- 54 Krokstad, S. *et al.* Cohort Profile: the HUNT Study, Norway. *Int J Epidemiol* **42**, 968-977, doi:10.1093/ije/dys095 (2013).
- 55 Holmen, J. *et al.* The Nord-Trøndelag Health Study 1995–97 (HUNT 2): objectives, contents, methods and participation. *Norsk epidemiologi* **13**, 19-32 (2003).
- 56 Bogduk, N. Management of chronic low back pain. *Medical journal of Australia* **180**, 79-83 (2004).
- 57 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283, doi:10.1038/ng.3643 (2016).
- 58 Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555-3557, doi:10.1093/bioinformatics/btv402 (2015).
- 59 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

- 60 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).
- 61 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 62 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
- 63 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 64 Smedley, D. *et al.* The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* **43**, W589-598, doi:10.1093/nar/gkv350 (2015).
- 65 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).
- 66 de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat Rev Genet* **17**, 353-364, doi:10.1038/nrg.2016.29 (2016).
- 67 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).
- 68 Gene\_Ontology\_Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res* **43**, D1049-1056, doi:10.1093/nar/gku1179 (2015).