

**A Genome-wide Association Study Identifies Novel Genetic Variants Associated  
with Neck or Shoulder Pain in the UK Biobank (N = 441,757)**

Yiwen Tao<sup>1,#</sup>, Qi Pan<sup>1,#</sup>, Tengda Cai<sup>1</sup>, Zen Haut Lu<sup>2</sup>, Mainul Haque<sup>3</sup>, Tania Dottorini<sup>4</sup>,  
Lesley A Colvin<sup>5</sup>, Blair H Smith<sup>5</sup>, Weihua Meng<sup>1,5,6\*</sup>.

1 Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute,  
University of Nottingham Ningbo China, Ningbo, China, 315100

2 PAPRSB Institute of Health Sciences, Universiti Brunei Darussalam, Bandar Seri  
Begawan, BE1410, Brunei Darussalam

3 School of Mathematical Sciences, University of Nottingham Ningbo China, Ningbo,  
China, 315100

4 School of Veterinary Medicine and Science, University of Nottingham, Nottingham, UK,  
LE12 5RD

5 Division of Population Health and Genomics, Ninewells Hospital and Medical School,  
University of Dundee, Dundee, UK, DD2 4BF

6 Center for Public Health, Faculty of Medicine, Health and Life Sciences, School of  
Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast, UK  
BT12 6BA

Yiwen Tao and Qi Pan contribute equally to the article

Corresponding Author: Dr Weihua Meng

Address: Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK, DD2 4BF; Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute, University of Nottingham Ningbo China, Ningbo, China, 315100. Tel: +44 1382 383419; +86 574 88188632. Email: [w.meng@dundee.ac.uk](mailto:w.meng@dundee.ac.uk); [weihua.meng@nottingham.edu.cn](mailto:weihua.meng@nottingham.edu.cn). ORCID: 0000-0001-5388-8494

#### ORCID codes

Yiwen Tao: 0009-0000-0102-4205

Qi Pan: 0000-0003-0135-9122

Tengda Cai: 0000-0003-1617-506X

Mainul Haque: 0000-0003-1836-8076

Tania Dottorini: 0000-0001-7733-906X

Lesley A Colvin: 0000-0002-1563-8600

Blair H Smith: 0000-0002-5362-9430

Weihua Meng: 0000-0001-5388-8494

#### **Abstract**

Neck and shoulder pain are prevalent musculoskeletal disorders that significantly impact the quality of life for a substantial portion of the global population. Studies have shown that women are more susceptible than men. This study aims to discover genetic variants associated with neck or shoulder pain through a genome-wide association study (GWAS), using data from 441,757 participants in the UK Biobank. The primary GWAS revealed five significant genetic loci (including two novel) associated with neck or shoulder pain, with the most significant single nucleotide polymorphism (SNP) being rs9889282 ( $p = 2.63 \times 10^{-12}$ ) near *CA10* on chromosome 17. Two novel significant associations were detected on

chromosomes 18 and 14, with the top SNPs being rs4608411 ( $p = 8.20 \times 10^{-9}$ ) near *TCF4* and rs370565192 ( $p = 3.80 \times 10^{-8}$ ) in *DCAF5*, respectively. The female-specific GWAS identified two significant loci including one near *CA10* and one near *LINC02770* on chromosome 1 with the top SNP being rs5779595 ( $p = 3.57 \times 10^{-8}$ ). The male-specific GWAS identified one locus in *SLC24A3* on chromosome 20 with the top SNP being rs16980973 ( $p = 6.52 \times 10^{-9}$ ). The tissue expression analysis revealed a significant association between brain tissues and neck or shoulder pain. In summary, this study has identified novel genetic variants for neck or shoulder pain. Sex stratified GWAS also suggested that gender played a role in the occurrence of the phenotype.

Keywords: Neck or shoulder pain; UK Biobank; genetic correlations; genome-wide association study; genomics; phenome-wide association analysis

## Introduction

Musculoskeletal pain is a serious concern worldwide, being the second leading cause of disability and with the fourth highest impact on overall health (Vos et al. 2012). Neck and shoulder pain are the second and third most prevalent musculoskeletal diseases, which are major causes of disability (Ferrari and Russell 2003; N. Sanchis et al. 2015). Given that the physical findings and symptoms of neck and shoulder discomfort are similar, these two types of pain are often discussed together (Fish et al. 2011; Meng et al. 2020). The characterization of neck and shoulder pain includes the appearance of stiffness, muscular tension, pressure, or pain in the areas stretching from the neck to the scapular arch (Iizuka et al. 2012).

There is a substantial quantity of epidemiological research indicating that the incidence and prevalence of neck or shoulder pain are high, however this varies widely between studies and nations. The prevalence of neck or shoulder pain in the general health examination program in Norway was 15.4% in men and 24.9% in women, with a sample size of 29,026 (Hasvold and Johnsen 1993). In 2019, the Philippines had one of the highest burdens of neck pain around the world, with an age-standardized point prevalence of 5,333.5 per 100,000 population, and an annual incidence rate of 1,156.3 per 100,000 population (Shin et al. 2022). From a global perspective, the age-standardized median prevalence of neck pain and shoulder pain was 27.0 per 1,000 and 37.8 per 1,000, respectively (Kazeminasab et al. 2022; Lucas et al. 2022). However, because the majority of people do not seek treatment for their problem, it is critical to assess the prevalence of shoulder discomfort in order to examine its influence on the population (Pribicevic 2012). For shoulder problems alone, it has been reported that approximately half of all new episodes that present in medical practice resolve completely within six months, rising to 60% after a year (van der Windt et al. 1996) In addition, frequent recurring neck discomfort not only has a significant influence on an individual's daily living activities and quality of life, but it also leads to lost productivity and ongoing economic expenditure, resulting in a significant social burden (Kanchanomai et al. 2011).

Several investigations have shown that the occurrence of neck or shoulder pain is significantly related to age, gender, body mass index, physical ability, psychological pressure, workplace and whether suffering from diabetes (Hviid Andersen et al. 2002; Hamberg-van Reenen et al. 2006; Iizuka et al. 2012; Hanvold et al. 2014; Pico-Espinosa

et al. 2017). Mahmud et al. (2012) reported that 72% of women and 51% of men suffer from neck pain, indicating that women are more likely to be affected by neck pain than men. Also, the prevalence of neck and shoulder pain in female adolescents increases with age (Hakala et al. 2002). In addition, compared to normal-weight women and men, obese women and men had a 19% and 22% greater risk of neck or shoulder pain, respectively (Nieman 2012). Since neck or shoulder pain is affected by many factors and seriously affects an individual's physical and mental health, understanding the pathogenesis of neck or shoulder pain is crucial for improving targeted management approaches.

The importance of genetic factors in neck or shoulder pain has been demonstrated. Based on twin studies, environmental and genetic influences on the self-report of neck or shoulder pain have been found (Nyman et al. 2011). It was shown that heritability accounted for 68% of the variance in susceptibility to non-specific neck discomfort at the age of 11-12 years, whereas individual contextual factors accounted for 32% of the diversity in phenotype (Ståhl et al. 2013). Furthermore, the additive genetic effect was twice as strong for concurrent low back pain and neck and shoulder pain as it was for low back pain alone and neck and shoulder pain alone (Nyman et al. 2011). In a previous genome-wide association study (GWAS) using the UK Biobank, the *FOXP2*, *LINC01572*, and *CA10* genes were identified to be significantly associated with neck or shoulder pain, with only *FOXP2* and *LINC01572* being marginally replicated in another cohort (Meng et al. 2020).

The purpose of the study was to identify novel genetic variants associated with neck or shoulder pain in the UK Biobank by conducting a GWAS on the phenotype using new

definitions. Compared with our previous GWAS by Meng et al.(2020), we used a new definition of controls which generated a larger sample size. In addition, we performed new sex-specific GWAS to explore potential genetic variants specifically associated with males or females. We also carried out post GWAS analyses which were not available at the time of the previous publication. Through this study, it was hoped that new genetic associations with neck or shoulder pain could be found, suggesting novel related genetic mechanisms and new targets for the treatment of neck or shoulder pain.

## **Methods**

### **Information about cohorts**

The UK BioBank is a unique cohort, facilitating research into health and disease. The cohort is a large-scale collection of broad environmental, lifestyle and genetic data from more than 500,000 volunteers aged 40 to 69 years in the UK from 2006 to 2010. The volunteers gave informed consent to complete detailed questionnaires and provide biological samples including urine, saliva and blood (available at [www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk)). Ethical approval for the study was obtained from the National Health Service and the National Research Ethics Service (reference 11/NW/0382). The genetic data and related data of half a million people in the UK are available for approved research projects into a wide range of diseases.

DNA extraction and quality control (QC) were standardized prior to the release of all data to the UK Biobank, and is described at [https://biobank.ctsu.ox.ac.uk/crystal/ukb/docs/genotyping\\_sample\\_workflow.pdf](https://biobank.ctsu.ox.ac.uk/crystal/ukb/docs/genotyping_sample_workflow.pdf). The Wellcome Trust Centre for Human Genetics at

Oxford University is responsible for ensuring the reliability of genotyping results. QC steps include identifying poorly performing markers, assessing sample correlations, and accounting for batch effects. These detailed QC steps are available at <http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580>.

### **Case and control definitions**

A questionnaire was designed by the UK Biobank including a specific pain-related question used in this study: “in the previous month, have you encountered any of the following that interfered with your regular activities?” The alternatives were as follows: (1) Headache; (2) Facial pain; (3) Neck or shoulder pain; (4) Back pain; (5) Neck or shoulder pain; (6) Hip pain; (7) Knee pain; (8) Pain all over the body; (9) None of the above; and (10) Prefer not to say (UK Biobank Questionnaire field ID: 6159). It is possible for volunteers to choose one or more answers. In this study, the cases of neck or shoulder pain were those who answered “yes” to “Neck or shoulder pain”. The control group was defined by those who did not choose both “Neck or shoulder pain” and “Prefer not to say”. In particular, this study used data from white British descent (field ID: 21000) in order to narrow the population stratification and make the results more specific and interpretable.

### **Design of the GWAS**

In the primary GWAS, we aimed to understand the genetic basis of neck or shoulder pain. Recognizing the significance of gender in the manifestation of health conditions, our secondary analysis focused on conducting two sex-stratified separate GWAS. This targeted approach in the two sex-specific GWAS aimed to unveil any distinct genetic signatures linked to neck or shoulder pain in males and females.

## **GWAS and statistical analysis**

Genome-wide complex trait analysis (GCTA, v1.94.1) is a statistical software for estimating the genetic contribution of complex traits across the genome, and is described at <https://yanglab.westlake.edu.cn/software/gcta/#Overview> (Yang et al. 2011). Utilizing the fastGWA function within GCTA, an ultra-fast mixed linear model association tool was performed for GWAS association analysis by using a sparse genetic relationship matrix. The QC step involved excluding single nucleotide polymorphisms (SNPs) with INFO scores below 0.3, minor allele frequencies lower than 0.5%, and failed Hardy-Weinberg tests ( $p < 10^{-6}$ ). In addition, mitochondrial SNPs and SNPs on the sex chromosomes were excluded. The association tests were adjusted for sex, age, BMI, and eight population principal components using the fastGWAS function in the primary GWAS. Similarly, the association tests were performed with adjustments for age, BMI, and eight population principal components in the secondary GWAS. R v4.2.2 was used to select the data of white British descent in UK Biobank data and divide the cases and controls of neck or shoulder pain. The difference in gender frequency between cases and controls was assessed by a  $\chi^2$  test and other covariates were tested by independent t-tests using R v4.2.2, with significance determined by a p-value threshold of 0.05. A threshold of  $p < 5 \times 10^{-8}$  is generally considered to be significant across the genome-wide association, which was applied for this study. Furthermore, GCTA was employed to calculate the narrow-sense heritability.

## **GWAS-associated analysis by FUMA**

The functional mapping and annotation (FUMA) of GWAS web application is an



annotation tool for gene analysis, gene-set analysis and tissue expression analysis of GWAS results (Watanabe et al. 2017). The SNP2GENE function accepts as input GWAS summary results and gives thorough functional annotation for all SNPs in genomic regions highlighted by lead SNPs. The number of related loci is ascertained using the "maximum distance of linkage disequilibrium (LD) blocks to merge" method of FUMA, while the number of significant independent SNPs is determined using the  $r^2$  value ( $r^2 > 0.6$ ). In detail, the default FUMA parameters were utilized including the threshold of significant SNPs ( $p < 5 \times 10^{-8}$ ) and the maximum distance of LD ( $r^2 > 0.6$ ) with a minimum minor allele frequency (MAF)  $> 0.01$ . Also, the maximum distance between LD blocks that might merge into a locus was set to less than 250 kb, and the 1000G Phase3 EUR population was utilized as the reference panel population. Furthermore, Locus Zoom (<http://locuszoom.org/>) was used to provide regional visualization (Pruim et al. 2010).

The three key aspects of analysis facilitated by FUMA include gene-based association analysis and gene-set analysis conducted by integrating MAGMA (v1.0619) and tissue expression analysis generated by GTEx. In the aspect of gene analysis, the summary statistics of SNPs were integrated at the level of whole genes. Statistical information about the associated SNPs is aggregated at each locus to obtain information about the association with the studied phenotype at the gene level. Integration of the statistical information of SNPs presents the degree of association between each gene and the phenotype under study. All SNPs found inside genes were specifically mapped to 19,023 protein-coding genes and the definition of the significance was  $p = 0.05/19,023 = 2.60 \times 10^{-6}$ . In the gene-set analysis, FUMA performs collective analysis of genomes that share

common biological, functional, or other features and the definition of the significance was set at  $p < 0.05/15,485 = 3.23 \times 10^{-6}$ . This approach helps to identify specific biological pathways, cellular functions, or other functional groups that further explain the genetic basis of the observed phenotypes. In the tissue expression analysis, FUMA uses tissue-specific expression data from GTEx to analyze the expression level of the specific genes in different tissues.

### **Genetic correlation analysis by LDSC**

Linkage disequilibrium score regression (LDSC) is a statistical method used to estimate heritability and genetic associations from GWAS summary statistics, which is described at <https://github.com/bulik/ldsc> (Bulik-Sullivan et al. 2015). LDSC reveals the association between different phenotypes by quantifying the LD between pairs of loci. In addition, LDSC can provide a genome-wide map of genetic correlation, showing the level of genetic correlation of different loci pairs, which helps to identify clusters of loci that are associated with the disease or phenotype as a whole. In this study, LDSC was also used to analyze the genetic correlation of neck or shoulder pain between males and females. The analysis was intended to discover overall genetic difference between male and female patients. We also calculated genetic connections of neck or shoulder discomfort using 1,396 characteristics from the UK Biobank via the Complex Trait Virtual Lab (CTG-VL) (<https://genoma.io/>). CTG-VL is a complementary and open-source program, which combines available GWAS datasets to facilitate genetic correlation estimation of complex characteristics using LDSC. The results were corrected for multiple testing using the Bonferroni method.

## **Phenome-wide association analysis (PheWAS)**

We conducted a Phenome-Wide Association Analysis (PheWAS) to explore associations between significant SNP associations and their corresponding genes with other traits with two objectives: (1), aiming to validate GWAS results by confirming the associations we identified with pain phenotypes; and (2) identify novel relationships between significant genetic variants associated with neck or shoulder pain and other phenotypes (focusing on psychiatric phenotypes such as morningness schizophrenia defined by ATLAS). This analysis involved generating PheWAS plots using an extensive dataset of 4,756 GWAS summary statistics available on the GWAS ATLAS platform (Watanabe et al. 2019). For this PheWAS, only SNPs with associations whose  $p$  values were lower than 0.05 were considered and adjustment for multiple comparisons was carried out using the Bonferroni correction method.

## **Results**

### **Description of the samples**

In the initial phase of the UK Biobank study (2006-2010), a cohort of 501,708 participants completed a pain questionnaire. Of these participants, 101,305 reported experiencing 'neck or shoulder pain', and were thus classified as 'cases' for the study. In contrast, 340,452 participants, who did not report such pain, were classified as 'controls'. Following the selection of white-British participants and the exclusion of samples failing to meet QC criteria, 98,652 cases (42,869 males and 55,783 females) were included, with 331,541 controls (153,827 males and 177,714 females) in the primary GWAS analysis. This

GWAS included 11,165,459 SNPs in the primary analysis. The secondary phase of our study involved sex-stratified GWAS analyses. In this phase, following identical QC procedures, the female cohort (233,497 samples) consisted of 55,784 cases and 177,713 controls, while the male cohort (196,696 samples) included 42,868 cases and 153,828 controls. Table 1 presents a comprehensive overview of the clinical attributes of both the cases and control of the primary and secondary GWAS.

### **GWAS Results**

In the primary GWAS, five GWAS signals were identified that demonstrated GWAS significant associations with neck or shoulder pain. These signals reached genome-wide significance with  $p$  values less than  $5 \times 10^{-8}$ , as shown in Figure 1. Of these five loci, two are newly identified. Table 2 offers a comprehensive overview of the top SNPs within each loci. For a detailed list of all SNPs that displayed significant associations in this GWAS, please refer to Supplementary Table 1.

The primary GWAS revealed the strongest association in the SNP cluster near the *CA10* gene on chromosome 17q21.33, a  $p$  value of  $2.63 \times 10^{-12}$  for rs9889282. A significant association was also identified in the *FOXP2* gene on chromosome 7, with a lowest  $p$  value of  $1.69 \times 10^{-9}$  for rs34291892. Another notable association was identified in the *SLC39A8* gene on chromosome 4, with a lowest  $p$  value of  $3.84 \times 10^{-9}$  for rs13135092. Additionally, significant associations were detected on chromosomes 18 and 14, with the top SNPs being rs4608411 ( $p = 8.20 \times 10^{-9}$ ) near *TCF4* and rs370565192 ( $p = 3.80 \times 10^{-8}$ ) in *DCAF5*, respectively. These loci on chromosomes 18 and 14 are reported here for the

first time in association with neck or shoulder pain. The regional plots displaying the most significant locus along with two novel loci are featured in Figures 2 and 3, while the regional plots for the additional two loci are available in Supplementary Figures 1. The Q–Q plot of the GWAS during the discovery phase is depicted in Supplementary Figure 2. The SNP-based heritability for neck or shoulder pain was determined to be 0.05, with a standard error of 0.02.

In the secondary GWAS analyses, which were stratified by sex, the male-specific GWAS identified a single locus of genome-wide significance associated with neck or shoulder pain. This locus, distinct from those found in the primary GWAS, was located in the *SLC24A3* gene on chromosome 20, with a lowest  $p$  value of  $6.52 \times 10^{-9}$  for rs16980973. The female-specific GWAS revealed two significant loci associated with neck or shoulder pain, one of which differed from the primary GWAS findings. This novel locus was identified near the *LINC02770* gene on chromosome 1, with a  $p$  value of  $3.57 \times 10^{-8}$  for rs5779595. Detailed information on these findings is available in Table 3, and the corresponding Manhattan plots can be found in Figure 1.

### **Gene, gene-set and tissue expression analysis by FUMA**

In our gene analysis of primary GWAS, *FOXP2* displayed the most significant association, with a  $p$  value of  $1.69 \times 10^{-9}$ . A total of six genes, namely *PABPC4*, *TCTA*, *MICB*, *FOXP2*, *IST1* and *SLC44A2*, were found to be associated with neck or shoulder pain, each with a  $p$  value less than  $2.60 \times 10^{-6}$  (0.05/19203). The detailed results of this analysis can be found in Supplementary Table 2.

In the gene-set analysis using MEGMA, a comprehensive examination of 15,485 gene sets was conducted. Of these, one gene set, specifically 'GOCC PRESYNAPTIC ENDOCYTIC ZONE', was identified as significantly associated ( $p = 2.95 \times 10^{-7}$ ). The top ten gene sets from this analysis have been detailed in Supplementary Table 3.

In the tissue expression analysis, the broader categories of 'Brain' and 'Pituitary' showed statistical significance among 30 general tissue types across various organs. Notably, eight brain-associated tissues, including 'Brain Cortex,' 'Brain Frontal Cortex BA9', 'Brain Anterior Cingulate Cortex BA24', 'Brain Caudate Basal Ganglia', 'Brain Nucleus Accumbens Basal Ganglia', 'Brain Hypothalamus', 'Brain Hippocampus' and 'Brain Amygdala' exhibited significant associations ( $p < 0.001$ ) out of 53 specific tissue types from different organs. Further details and visualizations of these findings can be found in Figures 4 and 5.

### **Genetic correlation analysis**

In our analysis of the genetic correlations between neck or shoulder pain and other phenotypes, several significant associations were identified. Neck or shoulder pain showed substantial genetic correlations with some other pain phenotypes, including multisite chronic pain ( $r_g = 0.89$ ,  $p = 0$ ), back pain ( $r_g = 0.82$ ,  $p = 1.06 \times 10^{-175}$ ), and hip pain ( $r_g = 0.77$ ,  $p = 9.26 \times 10^{-91}$ ). Additionally, significant positive genetic correlations were observed with neck or shoulder pain in relation to various medical conditions and health outcomes, including painful gums ( $r_g = 0.72$ ,  $p = 2.81 \times 10^{-23}$ ), usage of amitriptyline ( $r_g = 0.74$ ,  $p = 1.26 \times 10^{-16}$ ), and usage of co-codamol ( $r_g = 0.76$ ,  $p = 6.09 \times$

$10^{-38}$ ). These results are comprehensively documented in Supplementary Table 4 and depicted in Figure 6. The genetic correlation for neck or shoulder pain between males and females was also calculated ( $r_g = 0.79$ ,  $p = 2.67 \times 10^{-41}$ ), potentially providing extra evidence for the observed genetic differences in the sex-stratified secondary GWAS.

## **PheWAS**

PheWAS was conducted using the GWAS ATLAS platform to investigate phenotypes associated with independent significantly associated SNPs (rs9889282, rs34291892, rs13135092, rs4608411, rs370565192, rs12951067), and significantly associated genes (*SLC39A8*, *FOXP2*, *DCAF5*, *CA10*, and *TCF4*). rs9889282 was significantly associated with the self-rated health trait ( $p = 3.82 \times 10^{-7}$ ), while rs34291892 demonstrated a significant association with educational attainment ( $p = 1.70 \times 10^{-6}$ ). Both rs13135092 and rs4608411 were associated with schizophrenia ( $p = 3.60 \times 10^{-12}$  and  $p = 7.42 \times 10^{-5}$ , respectively). rs370565192 was associated with the Intelligence trait ( $p = 3.87 \times 10^{-5}$ ), and rs12951067 was associated with the frequency of tiredness/lethargy in the last 2 weeks ( $p = 4.25 \times 10^{-5}$ ). The genes *SLC39A8*, *FOXP2*, and *DCAF5* demonstrated associations with many traits following Bonferroni correction, covering a range of domains including ophthalmological, cellular, psychiatric, immunological, cardiovascular, and metabolic areas. The *CA10* and *TCF4* genes exhibited a strong association with psychiatric traits, specifically 'morningness' ( $p = 5.42 \times 10^{-11}$  for *CA10*) and schizophrenia ( $p = 3.64 \times 10^{-20}$  for *TCF4*). All significant traits that passed Bonferroni correction, are listed in

Supplementary Table 5. The plots of phenotypes associated with SNPs and genes are visualized in Supplementary Figures 3 and 4, respectively.

## Discussion

These three GWAS on neck or shoulder pain utilizing the UK Biobank dataset identified several notable genetic loci associated with the phenotype, including five loci (two novel) in the primary analysis, two loci in females and one locus in males (Table 2 and 3).

In the primary GWAS, the locus near *CA10* on chromosome 17 was the most significantly associated with neck or shoulder pain, with a top SNP of rs9889282 ( $p = 2.63 \times 10^{-12}$ ).

*CA10* encodes a protein from the zinc metalloenzyme family called carbonic anhydrase, involved in bone resorption and bone mineral solubilization (Okamoto et al. 2001; Pastorekova et al. 2004). The protein is also believed to be involved in the development of the brain and the central nervous system (<https://www.ncbi.nlm.nih.gov/gene/105371829#gene-expression>). The strong interconnections between carbonic anhydrase activity and neuropathic pain have been proved by Asiedu et al.(2010, 2014). Da Vitoria Lobo et al. (2022) concluded through experiments that inhibition of carbonic anhydrase activity by intraperitoneal injection of acetazolamide decreased neuropathic pain. For *CA10*, one study suggested that it has at least two global functions that do not interact (Sterky et al. 2017). First, overexpression of *CA10* results in increased surface levels, altered posttranslational modification patterns, and total levels of neuroxin proteins (Sterky et al. 2017). Secondly, *CA10* may play the role of adaptors, facilitating neurexins' indirect connections with unidentified postsynaptic target molecules, and thus mediating the



development of new transsynaptic complexes (Sterky et al. 2017). Evidence suggests that the mechanism by which regulatory proteins interact with neurexins-neuroligins to jointly form intersynaptic bridges plays a role in pathological pain (Li et al. 2022). As there are not many studies on *CA10* in humans, future studies should focus on elucidating the mechanism of *CA10* action on neck or shoulder pain.

The second most significantly associated locus was identified in the *FOXP2* gene located on chromosome 7, with a top SNP of rs34291892 ( $p = 1.69 \times 10^{-9}$ ). *FOXP2* is one member of the forkhead family of transcription factors and is located on human chromosome 7q31, which encodes a 715-amino acid protein (Enard et al. 2002). It comprises a glutamine-rich area made up of two neighboring polyglutamine tracts encoded by CAG and CAA repeat mixes, which are known to have high mutation rates (Enard et al. 2002). The *FOXP2* gene may play a significant role in the establishment of a putative frontostriatal network, which is associated with the function of learning, planning, and execution in the speech motor sequences, similar to that engaged in other types of motor abilities (Liégeois et al. 2003). In the avian striatum, *FOXP2* expression is controlled by singing, suggesting the prospect that developmental verbal dyspraxia resulting from human *FOXP2* mutation may largely represent a loss in ongoing brain signaling, rather than developmental miswiring (Deshpande and Lints 2013). Alternatively, genes *DLX5* and *SYT4*, which are affected by *FOXP2*, are essential for brain development and function (Meng et al. 2020). In a previous GWAS analysis, *FOXP2* has been confirmed to be associated with neck or shoulder pain, and multi-site chronic pain (Johnston et al. 2019; Meng et al. 2020). In addition, *FOXP2* has been found to be associated with the

development of attention deficit hyperactivity disorder and psychological disorders (Wendt et al. 2018; Meyer et al. 2023). The identified brain regions, including the hypothalamus, are known to play crucial roles in central neuroendocrine and hormone secretion (Markakis 2002). Studies of post-mortem brain microscopy have shown a link between the hypothalamus and mental illness (Bernstein et al. 2019). This connection raises questions about the psychological aspects influencing the development and persistence of neck or shoulder pain, which is consistent with the function.

The third notable association was identified in the *SLC39A8* gene on chromosome 4. *SLC39A8* encodes transporter zinc- and iron-related protein eight (ZIP8), which carry essential-metal divalent cations including zinc, copper, iron, and manganese (Gálvez-Peralta et al. 2012). In a previous study, foetal and neonatal mice with *SLC39A8* homozygote show reduced numbers of hematopoietic islands in the hematocrit, serum iron, red cell count, yolk sac and liver, low hemoglobin levels, and total iron-binding capacity—all consistent with the presence of severe anemia (Nebert and Liu 2019). GWAS subsequently revealed that human *SLC39A8*-deficient variants exhibit distinct multiorientation defects associated with clinical diseases in almost all organs including many developmental and congenital diseases, musculoskeletal system, central nervous system, coagulation system, cardiovascular system and immune system (Nebert and Liu 2019). In the details of the central nervous system, target validation tests demonstrated that microRNA 488 can target validation tests demonstrated that microRNA 488 can target *SLC39A8* mRNA, and inhibiting *SLC39A8* expression in the osteoarthritis animal model reduced cartilage deterioration (Song et al. 2013). A subsequent paper has suggested that

the increased expression of *SLC39A8* is associated with increased intracellular Zn levels in diseased chondrocytes among all Zn transporters in human and mouse cartilage affected by osteoarthritis (Kim et al. 2014). Musculoskeletal conditions, including osteoarthritis, can contribute to pain and discomfort in the neck and shoulder regions, making *SLC39A8* a potential player in these processes. The positive correlation between *SLC39A8* and osteoarthritis in the experiment with human and mouse embryos corroborates this conjecture. It is worth noting that this locus near *SLC39A8* has already been found to be associated with chronic neck or shoulder pain in a GWAS analysis of chronic musculoskeletal pain (Tsepilov et al. 2020).

The two novel loci are near *TCF4* on chromosome 18 with a top SNP of rs4608411 ( $p = 8.20 \times 10^{-9}$ ) and in *DCAF5* on chromosome 14 with a top SNP of rs370565192 ( $p = 3.80 \times 10^{-8}$ ), separately. The discovery of these two new loci represents a significant further step forward in unraveling the complex genetic architecture of neck or shoulder pain.

*TCF4* encodes transcription factor 4, a fundamental helix-loop-helix transcription factor. The encoded protein is a member of the ubiquitous E-protein family, causes severe CNS and autonomous nervous system dysfunction when mutated in humans (Amiel et al. 2007). Rare variants of *TCF4* including nonsense, missense, or microdeletion mutations may cause haploidy deficiency, which leads to a syndromic encephalopathy called Pitt-Hopkins syndrome (Brockschmidt et al. 2007; de Pontual et al. 2009; Forrest et al. 2012; Sweatt 2013; Li et al. 2019). In certain circumstances, it is harmful of the large amount of mutations missense mutations or create shortened proteins to the bHLH-binding domain crucial for the functional transcription activity of *TCF4* (Sepp et al. 2012; Sweatt 2013; Li et

al. 2019). Mutations in *TCF4* have also been linked to a number of other psychiatric disorders including autism, bipolar disorder and schizophrenia (Stefansson et al. 2009; Steinberg et al. 2011; Forrest et al. 2014; Hamdan et al. 2014). Meanwhile, *TCF4* shows expression in the adult brain, skeletal muscle, lung, and heart, confirming the disease's postnatal growth retardation and microcephaly (Amiel et al. 2007).

The *DCAF5* gene is one part of the ubiquitin ligase complex *DDB1-CUL4*, and the *DCAF5* gene's brain expression results in *DDB1*-and-*CUL4* related factor 5 (Angers et al. 2006; Lee and Zhou 2007; Śmigiel et al. 2020). *CUL4B* makes up the majority of this complex, and mutations in it can result in X-linked intellectual impairment (Tarpey et al. 2007). Likely haploinsufficiency of genes including *DCAF5* deletion are related to congenital heart defects, short fingers, and mild intellectual disability (Oehl-Jaschkowitz et al. 2014). It has been proposed that dysfunction of its binding partner *DCAF5* caused by de novo mutations may have similar consequences for developmental function (Śmigiel et al. 2020). *DCAF5* carries rare segregating variants in at least two independent families with bipolar disorder, which are hypothesized to be associated with psychiatric disorders (Forstner et al. 2020). In addition, a GWAS related to lung function identified rare functional coding variants in *DCAF5* that prevented rapid lung function decline (Mathias et al. 2012).

In the sex-stratified GWAS analyses, male-specific GWAS revealed a locus on *SLC24A3* on chromosome 20, with a top SNP of rs16980973 ( $p = 6.52 \times 10^{-9}$ ). *SLC24A3* encodes NCKX3, a potassium-dependent  $\text{Na}^+/\text{K}^+/\text{Ca}^{2+}$  exchanger that regulates intracellular calcium homeostasis (Citterio et al. 2011).  $\text{Ca}^{2+}$  has an important function in modulating

vascular tone, which contributes significantly to the regulation of systemic blood pressure (Karaki et al. 1997; Blaustein and Lederer 1999). Because there is not sufficient evidence to clarify the uniqueness of male-specific locus associated with neck or shoulder pain, more research should focus on the differences in neck or shoulder pain between males and females.

In female-specific GWAS analysis, the most significant variant was identified near *CA10* same as the primary GWAS. The other significant variant is near the *LINC02770* gene on chromosome 1 with the top SNP rs5779595 ( $p = 3.57 \times 10^{-8}$ ). In a genome-wide association study investigating associations with obstructive sleep apnea, *LINC02770* has been identified as being linked to this sleep disorder (Mohit et al. 2022). *LINC02770* is also thought to be associated with insomnia and post-traumatic stress disorder (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=LINC02770>). These findings align with clinical observations where individuals experiencing chronic neck or shoulder pain often report associated symptoms such as headaches, insomnia and mental illness including depression (Shahidi et al. 2015; Chu and Lee 2022). Although the specific genetic mechanism of *LINC02770* has not been explored, it introduces a novel aspect to the understanding of genetic factors contributing to pain susceptibility in females. The function of *LINC02770* and its potential involvement in pain modulation warrant further investigation to elucidate the intricate mechanisms. The genetic correlation for neck or shoulder pain between males and females was also calculated ( $r_g = 0.79$ ), which suggests that the genetic mechanisms of neck or shoulder pain are not the exactly same

between males and females. The reason for the difference is likely due to the important role of gender differences in the heterogeneity of pain, especially neuropathic pain (Gao et al. 2024; Smith 2024). These sex-specific genetic associations underscore the importance of considering biological differences between males and females in studies related to pain genetics. Tailoring interventions according to the identified genetic factors could lead to more effective and targeted pain management strategies, ultimately improving the quality of life for individuals suffering from these conditions.

Compared to the previous GWAS on neck or shoulder pain by Meng et al. (2020), this study discovered two new loci. In this study, the SNP-based heritability for neck or shoulder pain was determined to be 0.05, which is smaller than the previously estimate (0.115) by-Meng et al. (2020). The likely reason is the different definition of controls, which leads to a larger number of people in the control group. The scope of the control group in this study is larger than that of the former. Since the scope of the case group is unchanged, a larger number of controls is likely to lead to a decrease in the SNP heritability of the GWAS result.

In the tissue expression analysis, neck or shoulder pain showed a strong correlation with the brain and pituitary categories. The identified brain tissues revealed key signaling pathways related to pain perception (BA24, hypothalamus, amygdala), sensory processing (amygdala, cortex, BA9, basal ganglia), and central neuroendocrine (hypothalamus) (Hudson 2000; Peyron et al. 2000; Markakis 2002; Kerestes et al. 2012; Lanciego et al. 2012). These findings align with clinical observations where individuals experiencing chronic neck or shoulder pain often report associated symptoms such as

headaches, fatigue, and alterations in mood (Shahidi et al. 2015; Chu and Lee 2022). The identified brain regions, including the cerebral cortex and amygdala, are known to play crucial roles in emotional processing and stress response (Šimić et al. 2021). This connection raises questions about the psychosocial aspects influencing the development and persistence of neck and shoulder pain. Moreover, tissue specific expression in the pituitary gland displayed a significant association with neck and shoulder pain. This finding suggests a potential involvement of hormonal regulation in the manifestation of pain in these regions, and could be at least part of the explanation for the sex differences. The intricate network between the brain, pituitary gland, and chronic pain in the neck or shoulder highlights the need for a comprehensive understanding of both neurological and endocrine factors contributing to this phenomenon.

The genetic correlations between neck or shoulder pain and several other pain phenotypes suggest that these disorders share common genetic determinants. The three most highly correlated phenotypes of neck or shoulder pain were multisite chronic pain ( $r_g = 0.89$ ), back pain ( $r_g = 0.82$ ), and hip pain ( $r_g = 0.77$ ). In addition, a correlation emerged between neck or shoulder pain and gum pain ( $r_g = 0.74$ ), highlighting the interconnectedness of these symptoms. Understanding these correlations could provide valuable insights into potential shared pathways or mechanisms underlying these health outcomes. Neck or shoulder pain was observed to be associated with various medical conditions and health outcomes. Moving beyond genetic correlations, the associations with specific medications add another layer of complexity. The correlation between neck or shoulder pain and receipt of amitriptyline ( $r_g = 0.74$ ) and co-codamol ( $r_g = 0.76$ )

suggests that biological links exist between the neck or shoulder pain and each phenotype as amitriptyline is an anti-depression drug although not used as a first-line treatment for depression due to considerable side effects and co-codamol is a pain-relief medicine. In the PheWAS, the results demonstrate that the top SNPs and corresponding genes are primarily highly associated with numerous psychiatric traits. This may provide further evidence that psychiatric factors could be an important contributor to neck or shoulder pain and/or vice versa. Meanwhile, these newly identified SNPs and genes are likely to play a role in the biological mechanism of neck or shoulder pain, which will have a suggestive effect on the development of targeted therapies (such as carbonic anhydrase inhibitors) for the pain. The shared genetic determinants, correlations with other pain phenotypes, and connections to various medical conditions and medications underscore the need for a holistic approach in understanding and addressing neck or shoulder pain.

It is important to note that cases and controls in the UK Biobank are self-reported and may be subject to several biases such as social expectations, recall periods, sampling methods or selective recall. Therefore, these potential effects need to be carefully considered when interpreting the findings. A potential limitation is that the control group contains persons with pain at other body sites. This may have an impact on the GWAS results. Genetic correlation studies revealed that many pain phenotypes are strongly genetically related. For example, as identified above (REF) there is a particularly strong genetic correlation between back pain and neck or shoulder pain. As a sensitivity analysis we repeated the above GWAS after excluding people who reported back pain in the control group. This re-analysis led to similar results in the genetic associations identified



(Supplementary Figure 5 and Supplementary Table 6). Additionally, we described cases and controls with neck or shoulder pain based on responses to specific questions from UK Biobank participants. However, this question was not designed to cover more detailed information about the severity, frequency, and exact location of the pain. Therefore, the phenotypic definitions we derived from these responses should be viewed as generalized. Since the sample size in this study was very large (about 450,000), it has sufficient study power to detect contributing genetic variants. We noticed that the  $p$  values of some loci associations were just smaller than the GWAS significance, therefore it is likely that the results can only be replicated by cohorts of a similar size as the UK Biobank, or larger. Neck or shoulder pain phenotypes were also not recorded by many other large publicly available genetic cohorts such as FinnGen (Kurki et al. 2023), and we therefore did not conduct replication analysis.

## **Conclusion**

In summary, our primary GWAS identified five genetic loci including two novel ones associated with neck or shoulder pain among all people with neck or shoulder pain. Our secondary GWAS identified a single novel genetic locus associated with neck or shoulder pain among males and two genetic loci (including one novel) associated with neck or shoulder pain among females. These discoveries not only expand our understanding of the genetic underpinnings of neck or shoulder pain but also emphasize the importance of considering sex-specific influences.

## **Funding**

This study was mainly funded by the Pioneer and Leading Goose R&D Program of Zhejiang Province 2023 with reference number 2023C04049 and Ningbo International Collaboration Program 2023 with reference number 2023H025. This work was also supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 633491 (DOLORisk).

### **Data Availability and Acknowledgement**

This study adheres to all ethical guidelines and data protection protocols of the UK Biobank. The current study was conducted under approved UK Biobank data application number 89386. The summary statistics of the UK Biobank results on neck or shoulder pain can be accessed upon publication. Any other data relevant to the study that are not included in the article or its supplementary materials are available from the authors upon reasonable request.

### **Author information**

#### **Authors and Affiliations**

**Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute,**

**University of Nottingham Ningbo China, Ningbo, China, 315100**

Yiwen Tao, Qi Pan, Tengda Cai, Weihua Meng

**PAPRSB Institute of Health Sciences, Universiti Brunei Darussalam, Bandar Seri**

**Begawan, BE1410, Brunei Darussalam**

Zen Haut Lu

**School of Mathematical Sciences, University of Nottingham Ningbo China, Ningbo,**

**China, 315100**

Mainul Haque

**School of Veterinary Medicine and Science, University of Nottingham, Nottingham,**

**UK, LE12 5RD**

Tania Dottorini

**Division of Population Health and Genomics, Ninewells Hospital and Medical**

**School, University of Dundee, Dundee, UK, DD2 4BF**

Weihua Meng, Lesley A Colvin, Blair H Smith

**Center for Public Health, Faculty of Medicine, Health and Life Sciences, School of**

**Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Belfast,**

**UK BT12 6BA**

Weihua Meng

### **Authors' Contributions**

All authors contributed to the analysis plan. YT and QP performed the UK Biobank GWAS analysis and drafted the paper. TC contributed to data formatting. ZHL, MH, TD, LAC and BHS provided comments to the paper. WM organised the project and provided comments.

### **Corresponding authors**

Correspondence to Weihua Meng.

### **Consent to Publish**

All authors provide consent for publication.

### **Ethical Approval**

This study was approved by the Ethics Committee of the University of Nottingham Ningbo China.

## References

- Amiel J, Rio M, Pontual L de, et al (2007) Mutations in TCF4, Encoding a Class I Basic Helix-Loop-Helix Transcription Factor, Are Responsible for Pitt-Hopkins Syndrome, a Severe Epileptic Encephalopathy Associated with Autonomic Dysfunction. *The American Journal of Human Genetics* 80:988–993. <https://doi.org/10.1086/515582>
- Angers S, Li T, Yi X, et al (2006) Molecular architecture and assembly of the DDB1–CUL4A ubiquitin ligase machinery. *Nature* 443:590–593. <https://doi.org/10.1038/nature05175>
- Asiedu M, Ossipov MH, Kaila K, Price TJ (2010) Acetazolamide and midazolam act synergistically to inhibit neuropathic pain. *PAIN* 148:302–308. <https://doi.org/10.1016/j.pain.2009.11.015>
- Asiedu MN, Mejia GL, Hübner CA, et al (2014) Inhibition of Carbonic Anhydrase Augments GABAA Receptor-Mediated Analgesia via a Spinal Mechanism of Action. *The Journal of Pain* 15:395–406. <https://doi.org/10.1016/j.jpain.2014.01.001>
- Bernstein H-G, Dobrowolny H, Bogerts B, et al (2019) The hypothalamus and neuropsychiatric disorders: psychiatry meets microscopy. *Cell Tissue Res* 375:243–258. <https://doi.org/10.1007/s00441-018-2849-3>

Blaustein MP, Lederer WJ (1999) Sodium/Calcium Exchange: Its Physiological Implications. *Physiological Reviews* 79:763–854.  
<https://doi.org/10.1152/physrev.1999.79.3.763>

Brockschmidt A, Todt U, Ryu S, et al (2007) Severe mental retardation with breathing abnormalities (Pitt–Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. *Human Molecular Genetics* 16:1488–1494. <https://doi.org/10.1093/hmg/ddm099>

Bulik-Sullivan BK, Loh P-R, Finucane HK, et al (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47:291–295. <https://doi.org/10.1038/ng.3211>

Chu EC-P, Lee LY-K (2022) Cervical Spondylosis as a Hidden Contributing Factor to Fibromyalgia: A Case Report. *Int Med Case Rep J* 15:639–646.  
<https://doi.org/10.2147/IMCRJ.S382872>

Citterio L, Simonini M, Zagato L, et al (2011) Genes Involved in Vasoconstriction and Vasodilation System Affect Salt-Sensitive Hypertension. *PLOS ONE* 6:e19620.  
<https://doi.org/10.1371/journal.pone.0019620>

Da Vitoria Lobo ME, Weir N, Hardowar L, et al (2022) Hypoxia-induced carbonic anhydrase mediated dorsal horn neuron activation and induction of neuropathic pain. *Pain* 163:2264–2279. <https://doi.org/10.1097/j.pain.0000000000002627>

de Pontual L, Mathieu Y, Golzio C, et al (2009) Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. *Human Mutation* 30:669–676.  
<https://doi.org/10.1002/humu.20935>

Deshpande M, Lints TJ (2013) The Molecular Convergence of Birdsong and Speech. In: Helekar SA (ed) *Animal Models of Speech and Language Disorders*. Springer, New York, NY, pp 109–184

Enard W, Przeworski M, Fisher SE, et al (2002) Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418:869–872.  
<https://doi.org/10.1038/nature01025>

Ferrari R, Russell AS (2003) Neck pain. *Best Practice & Research Clinical Rheumatology* 17:57–70. [https://doi.org/10.1016/S1521-6942\(02\)00097-9](https://doi.org/10.1016/S1521-6942(02)00097-9)

Fish DE, Gerstman BA, Lin V (2011) Evaluation of the Patient with Neck Versus Shoulder Pain. *Physical Medicine and Rehabilitation Clinics of North America* 22:395–410.  
<https://doi.org/10.1016/j.pmr.2011.03.009>

Forrest M, Chapman RM, Doyle AM, et al (2012) Functional analysis of TCF4 missense mutations that cause Pitt–Hopkins syndrome. *Human Mutation* 33:1676–1686.  
<https://doi.org/10.1002/humu.22160>

Forrest MP, Hill MJ, Quantock AJ, et al (2014) The emerging roles of TCF4 in disease and development. *Trends in Molecular Medicine* 20:322–331.

<https://doi.org/10.1016/j.molmed.2014.01.010>

Forstner AJ, Fischer SB, Schenk LM, et al (2020) Whole-exome sequencing of 81 individuals from 27 multiply affected bipolar disorder families. *Transl Psychiatry* 10:1–10. <https://doi.org/10.1038/s41398-020-0732-y>

Gálvez-Peralta M, He L, Jorge-Nebert LF, et al (2012) ZIP8 Zinc Transporter: Indispensable Role for Both Multiple-Organ Organogenesis and Hematopoiesis In Utero. *PLoS One* 7:e36055. <https://doi.org/10.1371/journal.pone.0036055>

Gao Q, Liu M-Q, Li J-X, et al (2024) Sex differences in stress-induced hyperalgesia and its mechanisms. *Journal of Neuroscience Research* 102:e25266. <https://doi.org/10.1002/jnr.25266>

Hakala P, Rimpelä A, Salminen JJ, et al (2002) Back, neck, and shoulder pain in Finnish adolescents: national cross sectional surveys. *BMJ* 325:743. <https://doi.org/10.1136/bmj.325.7367.743>

Hamberg-van Reenen HH, Ariëns GA, Blatter BM, et al (2006) Is an imbalance between physical capacity and exposure to work-related physical factors associated with low-back, neck or shoulder pain? *Scandinavian Journal of Work, Environment & Health* 32:190–197

Hamdan FF, Srour M, Capo-Chichi J-M, et al (2014) De Novo Mutations in Moderate or Severe Intellectual Disability. *PLOS Genetics* 10:e1004772.

<https://doi.org/10.1371/journal.pgen.1004772>

Hanvold TN, Wærsted M, Mengshoel AM, et al (2014) A longitudinal study on risk factors for neck and shoulder pain among young adults in the transition from technical school to working life. *Scandinavian Journal of Work, Environment & Health* 40:597–609. <https://doi.org/10.5271/sjweh.3437>

Hasvold T, Johnsen R (1993) Headache and neck or shoulder pain - frequent and disabling complaints in the general population. *Scandinavian Journal of Primary Health Care* 11:219–224. <https://doi.org/10.3109/02813439308994834>

Hudson AJ (2000) Pain Perception and Response: Central Nervous System Mechanisms. *Canadian Journal of Neurological Sciences* 27:2–16. <https://doi.org/10.1017/S0317167100051908>

Hviid Andersen J, Kaergaard A, Frost P, et al (2002) Physical, Psychosocial, and Individual Risk Factors for Neck/Shoulder Pain With Pressure Tenderness in the Muscles Among Workers Performing Monotonous, Repetitive Work. *Spine* 27:660

Iizuka Y, Shinozaki T, Kobayashi T, et al (2012) Characteristics of neck and shoulder pain (called katakori in Japanese) among members of the nursing staff. *Journal of Orthopaedic Science* 17:46–50. <https://doi.org/10.1007/s00776-011-0177-5>

Johnston KJA, Adams MJ, Nicholl BI, et al (2019) Genome-wide association study of multisite chronic pain in UK Biobank. *PLOS Genetics* 15:e1008164.



<https://doi.org/10.1371/journal.pgen.1008164>

Kanchanomai S, Janwantanakul P, Pensri P, Jiamjarasrangsi W (2011) Risk factors for the onset and persistence of neck pain in undergraduate students: 1-year prospective cohort study. *BMC Public Health* 11:566.  
<https://doi.org/10.1186/1471-2458-11-566>

Karaki H, Ozaki H, Hori M, et al (1997) Calcium movements, distribution, and functions in smooth muscle. *Pharmacol Rev* 49:157–230

Kazeminasab S, Nejadghaderi SA, Amiri P, et al (2022) Neck pain: global epidemiology, trends and risk factors. *BMC Musculoskeletal Disorders* 23:26.  
<https://doi.org/10.1186/s12891-021-04957-4>

Kerestes R, Ladouceur CD, Meda S, et al (2012) Abnormal prefrontal activity subserving attentional control of emotion in remitted depressed patients during a working memory task with emotional distracters. *Psychological Medicine* 42:29–40.  
<https://doi.org/10.1017/S0033291711001097>

Kim J-H, Jeon J, Shin M, et al (2014) Regulation of the catabolic cascade in osteoarthritis by the zinc-ZIP8-MTF1 axis. *Cell* 156:730–743.  
<https://doi.org/10.1016/j.cell.2014.01.007>

Kurki MI, Karjalainen J, Palta P, et al (2023) FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 613:508–518.

<https://doi.org/10.1038/s41586-022-05473-8>

Lanciego JL, Luquin N, Obeso JA (2012) Functional Neuroanatomy of the Basal Ganglia.

Cold Spring Harb Perspect Med 2:a009621.

<https://doi.org/10.1101/cshperspect.a009621>

Lee J, Zhou P (2007) DCAFs, the missing link of the CUL4-DDB1 ubiquitin ligase. Mol Cell

26:775–780. <https://doi.org/10.1016/j.molcel.2007.06.001>

Li H, Guo R, Guan Y, et al (2022) Modulation of Trans-Synaptic Neurexin–Neuroigin

Interaction in Pathological Pain. Cells 11:1940.

<https://doi.org/10.3390/cells11121940>

Li H, Zhu Y, Morozov YM, et al (2019) Disruption of TCF4 regulatory networks leads to

abnormal cortical development and mental disabilities. Mol Psychiatry

24:1235–1246. <https://doi.org/10.1038/s41380-019-0353-0>

Liégeois F, Baldeweg T, Connelly A, et al (2003) Language fMRI abnormalities associated

with FOXP2 gene mutation. Nat Neurosci 6:1230–1237.

<https://doi.org/10.1038/nn1138>

Lucas J, van Doorn P, Hegedus E, et al (2022) A systematic review of the global

prevalence and incidence of shoulder pain. BMC Musculoskelet Disord 23:1073.

<https://doi.org/10.1186/s12891-022-05973-8>

Mahmud N, Kenny DT, Rahman HA (2012) The effect of workplace office ergonomics

intervention on reducing neck and shoulder complaints and sickness absence.

International Proceedings of Economics Development & Research 42:142–9

Markakis EA (2002) Development of the neuroendocrine hypothalamus. *Frontiers in Neuroendocrinology* 23:257–291. [https://doi.org/10.1016/S0091-3022\(02\)00003-1](https://doi.org/10.1016/S0091-3022(02)00003-1)

Mathias RA, Kim Y, Taub M, et al (2012) Exome Sequencing Identifies Rare Functional Coding Variants In HK2, MTRR And DCAF5 That Determine Lung Function Decline In COPD. In: D50. GENETIC STUDIES IN COPD. American Thoracic Society, pp A5870–A5870

Meng W, Chan BW, Harris C, et al (2020) A genome-wide association study finds genetic variants associated with neck or shoulder pain in UK Biobank. *Human Molecular Genetics* 29:1396–1404. <https://doi.org/10.1093/hmg/ddaa058>

Meyer GP, da Silva BS, Bandeira CE, et al (2023) Dissecting the cross-trait effects of the FOXP2 GWAS hit on clinical and brain phenotypes in adults with ADHD. *Eur Arch Psychiatry Clin Neurosci* 273:15–24. <https://doi.org/10.1007/s00406-022-01388-7>

Mohit, Sharma I, Sharma V, et al (2022) Empirical assessment of allele frequencies of genome wide association study variants associated with obstructive sleep apnea. *Am J Transl Res* 14:3464–3471

N. Sanchis M, Lluch E, Nijs J, et al (2015) The role of central sensitization in shoulder pain: A systematic literature review. *Seminars in Arthritis and Rheumatism* 44:710–716.

<https://doi.org/10.1016/j.semarthrit.2014.11.002>

Nebert DW, Liu Z (2019) SLC39A8 gene encoding a metal ion transporter: discovery and bench to bedside. *Human Genomics* 13:51.  
<https://doi.org/10.1186/s40246-019-0233-3>

Nieman DC (2012) Physical Exercise, Body Mass Index, and Risk of Chronic Pain in the Low Back and Neck/Shoulders: Longitudinal Data From the Nord-Trøndelag Health Study. *Yearbook of Sports Medicine* 2012:1–2.  
<https://doi.org/10.1093/aje/kwr087>

Nyman T, Mulder M, Iliadou A, et al (2011) High Heritability for Concurrent Low Back and Neck-Shoulder Pain: A Study of Twins. *Spine* 36:E1469.  
<https://doi.org/10.1097/BRS.0b013e3181e2c878>

Oehl-Jaschkowitz B, Vanakker OM, De Paepe A, et al (2014) Deletions in 14q24.1q24.3 are associated with congenital heart defects, brachydactyly, and mild intellectual disability. *Am J Med Genet A* 164A:620–626.  
<https://doi.org/10.1002/ajmg.a.36321>

Okamoto N, Fujikawa-Adachi K, Nishimori I, et al (2001) cDNA sequence of human carbonic anhydrase-related protein, CA-RP X: mRNA expressions of CA-RP X and XI in human brain<sup>1</sup>The sequence reported in this paper has been deposited in the GenBank database under accession No. AB036836.1. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1518:311–316.

[https://doi.org/10.1016/S0167-4781\(01\)00193-2](https://doi.org/10.1016/S0167-4781(01)00193-2)

Pastorekova S, Parkkila S, Pastorek J, Supuran CT (2004) Carbonic anhydrases: current state of the art, therapeutic applications and future prospects. *J Enzyme Inhib Med Chem* 19:199–229. <https://doi.org/10.1080/14756360410001689540>

Peyron R, Laurent B, García-Larrea L (2000) Functional imaging of brain responses to pain. A review and meta-analysis (2000). *Neurophysiologie Clinique/Clinical Neurophysiology* 30:263–288. [https://doi.org/10.1016/S0987-7053\(00\)00227-6](https://doi.org/10.1016/S0987-7053(00)00227-6)

Pico-Espinosa OJ, Skillgate E, Tettamanti G, et al (2017) Diabetes mellitus and hyperlipidaemia as risk factors for frequent pain in the back, neck and/or shoulders/arms among adults in Stockholm 2006 to 2010 – Results from the Stockholm Public Health Cohort. *Scandinavian Journal of Pain* 15:1–7. <https://doi.org/10.1016/j.sjpain.2016.11.005>

Pribicevic M (2012) The Epidemiology of Shoulder Pain: A Narrative Review of the Literature. In: *Pain in Perspective*. IntechOpen, p 149

Pruim RJ, Welch RP, Sanna S, et al (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26:2336–2337. <https://doi.org/10.1093/bioinformatics/btq419>

Sepp M, Pruunsild P, Timmusk T (2012) Pitt–Hopkins syndrome-associated mutations in TCF4 lead to variable impairment of the transcription factor function ranging from

hypomorphic to dominant-negative effects. *Human Molecular Genetics* 21:2873–2888. <https://doi.org/10.1093/hmg/dds112>

Shahidi B, Curran-Everett D, Maluf KS (2015) Psychosocial, Physical, and Neurophysiological Risk Factors for Chronic Neck Pain: A Prospective Inception Cohort Study. *The Journal of Pain* 16:1288–1299. <https://doi.org/10.1016/j.jpain.2015.09.002>

Shin DW, Shin JI, Koyanagi A, et al (2022) Global, regional, and national neck pain burden in the general population, 1990–2019: An analysis of the global burden of disease study 2019. *Front Neurol* 13:955367. <https://doi.org/10.3389/fneur.2022.955367>

Šimić G, Tkalčić M, Vukić V, et al (2021) Understanding Emotions: Origins and Roles of the Amygdala. *Biomolecules* 11:823. <https://doi.org/10.3390/biom11060823>

Śmigiel R, Biela M, Szmyd K, et al (2020) Rapid Whole-Exome Sequencing as a Diagnostic Tool in a Neonatal/Pediatric Intensive Care Unit. *Journal of Clinical Medicine* 9:2220. <https://doi.org/10.3390/jcm9072220>

Smith PA (2024) BDNF in Neuropathic Pain; the Culprit that Cannot be Apprehended. *Neuroscience* 543:49–64. <https://doi.org/10.1016/j.neuroscience.2024.02.020>

Song J, Kim D, Lee CH, et al (2013) MicroRNA-488 regulates zinc transporter SLC39A8/ZIP8 during pathogenesis of osteoarthritis. *J Biomed Sci* 20:31.

<https://doi.org/10.1186/1423-0127-20-31>

Ståhl M k., El-Metwally A a., Mikkelsen M k., et al (2013) Genetic and environmental influences on non-specific neck pain in early adolescence: A classical twin study. *European Journal of Pain* 17:791–798. <https://doi.org/10.1002/j.1532-2149.2012.00247.x>

Stefansson H, Ophoff RA, Steinberg S, et al (2009) Common variants conferring risk of schizophrenia. *Nature* 460:744–747. <https://doi.org/10.1038/nature08186>

Steinberg S, de Jong S, Andreassen OA, et al (2011) Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Human Molecular Genetics* 20:4076–4081. <https://doi.org/10.1093/hmg/ddr325>

Sterky FH, Trotter JH, Lee S-J, et al (2017) Carbonic anhydrase-related protein CA10 is an evolutionarily conserved pan-neurexin ligand. *Proceedings of the National Academy of Sciences* 114:E1253–E1262. <https://doi.org/10.1073/pnas.1621321114>

Sweatt JD (2013) Pitt–Hopkins Syndrome: intellectual disability due to loss of TCF4-regulated gene transcription. *Exp Mol Med* 45:e21–e21. <https://doi.org/10.1038/emm.2013.32>

Tarpey PS, Raymond FL, O'Meara S, et al (2007) Mutations in CUL4B, which encodes a ubiquitin E3 ligase subunit, cause an X-linked mental retardation syndrome

associated with aggressive outbursts, seizures, relative macrocephaly, central obesity, hypogonadism, pes cavus, and tremor. *Am J Hum Genet* 80:345–352.  
<https://doi.org/10.1086/511134>

Tsepilov YA, Freidin MB, Shadrina AS, et al (2020) Analysis of genetically independent phenotypes identifies shared genetic factors associated with chronic musculoskeletal pain conditions. *Commun Biol* 3:1–13.  
<https://doi.org/10.1038/s42003-020-1051-9>

van der Windt DA, Koes BW, Boeke AJ, et al (1996) Shoulder disorders in general practice: prognostic indicators of outcome. *Br J Gen Pract* 46:519–523

Vos T, Flaxman AD, Naghavi M, et al (2012) Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* 380:2163–2196.  
[https://doi.org/10.1016/S0140-6736\(12\)61729-2](https://doi.org/10.1016/S0140-6736(12)61729-2)

Watanabe K, Stringer S, Frei O, et al (2019) A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* 51:1339–1348.  
<https://doi.org/10.1038/s41588-019-0481-0>

Watanabe K, Taskesen E, van Bochoven A, Posthuma D (2017) Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 8:1826.  
<https://doi.org/10.1038/s41467-017-01261-5>



Wendt FR, Carvalho CM, Gelernter J, Polimanti R (2018) DRD2 and FOXP2 are implicated in the associations between computerized device use and psychiatric disorders. 497420

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet 88:76–82.  
<https://doi.org/10.1016/j.ajhg.2010.11.011>

**Figure legends:**

**Fig. 1** The Manhattan plot of the primary GWAS analysis on neck or shoulder pain (N = 441,757), the female-specific GWAS analysis (N = 239,772) and the male-specific GWAS analysis (N= 201,985)

The dashed red line indicates the cut-off  $p$  value of  $5 \times 10^{-8}$

**Fig. 2** The regional plots of locus in the CA10 region

**Fig. 3** The regional plots of locus in the TCF4 and DCAF5 regions

**Fig. 4** Tissue expression results on 30 specific tissue types by GTEx in the FUMA

The dashed line shows the cut-off  $p$  value for significance with Bonferroni adjustment for multiple hypothesis testing

**Fig. 5** Tissue expression results on 53 specific tissue types by GTEx in the FUMA

The dashed line shows the cut-off  $p$  value for significance with Bonferroni adjustment for multiple hypothesis testing

**Fig.6** Genetic correlation results for pain intensity using LDSC on CTG-VL. For a full list, see Supplementary Table 4

**Table 1** Clinical characteristics of neck or shoulder pain cases and controls in the UK Biobank

GWAS Analysis	Covariates	Cases	Controls	p-value
Primary GWAS	Sex (male: female)	42,869:55,783	153,827:177,714	<0.001
	Age (years)	56.8 (7.95)	56.9 (8.01)	<0.001
	BMI (kg/m <sup>3</sup> )	27.9 (5.00)	27.3 (4.70)	<0.001
Male-specific GWAS	Age (years)	57.1 (8.06)	57.0 (8.09)	<0.05
	BMI (kg/m <sup>3</sup> )	28.3 (4.42)	27.7 (4.19)	<0.001
Female-specific GWAS	Age (years)	56.7 (7.85)	56.6 (7.94)	<0.05
	BMI (kg/m <sup>3</sup> )	27.6 (5.39)	26.9 (5.07)	<0.001

Total number of females: 233,497; Total number of males: 196,696.

A  $\chi^2$  test was used to test the difference of gender frequency between cases and controls and an independent t test was used for other covariates.

Continuous covariates were presented as mean (standard deviation)

BMI represents body mass index.

**Table 2** The top SNPs within 5 loci identified by the GWAS on neck or shoulder pain

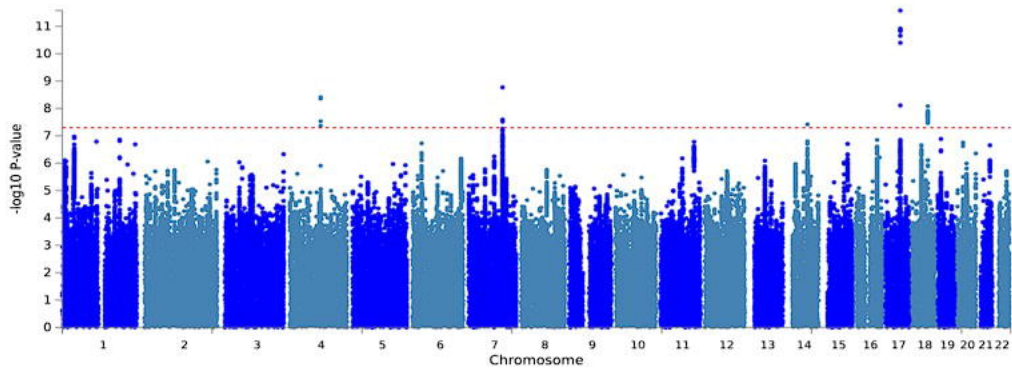
Locus Rank	rsID	Chr	SNP position	Nearest Gene	Effect allele	Non effective allele	Frequency the effect allele	<i>p</i> value	Beta	Identified or novel
1	rs9889282	17	50259142	<i>CA10</i>	A	C	0.61	$2.63 \times 10^{-12}$	-0.007	Meng et al.
2	rs34291892	7	114058731	<i>FOXP2</i>	CA	C	0.62	$1.69 \times 10^{-9}$	0.005	Meng et al.
3	rs13135092	4	103198082	<i>SLC39A8</i>	A	G	0.92	$3.84 \times 10^{-9}$	-0.01	Tsepilov et al.
4	rs4608411	18	52824637	<i>TCF4</i>	C	T	0.67	$8.20 \times 10^{-9}$	-0.005	novel
5	rs370565192	14	69515858	<i>DCAF5</i>	TC	T	0.63	$3.80 \times 10^{-8}$	-0.005	novel

Chr: chromosome

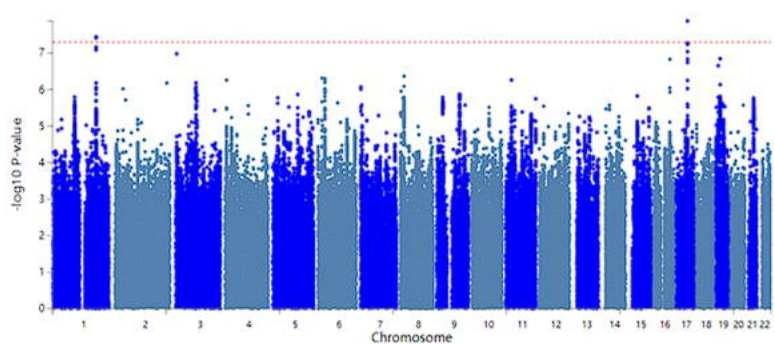
**Table 3** The top SNPs within 3 loci identified by the male-specific GWAS and female-specific GWAS on neck or shoulder pain

Secondary GWAS	Locus Rank	rsID	Chr	SNP position	Nearest Gene	Effect allele	Non effective allele	Frequency the effect allele	p value	Beta
Male-specific GWAS	1	rs16980973	20	19648493	<i>SLC24A3</i>	A	T	0.87	$6.52 \times 10^{-9}$	0.012
Female-specific GWAS	1	rs9889282	17	50259142	<i>CA10</i>	A	C	0.61	$1.31 \times 10^{-8}$	-0.007
	2	rs5779595	1	191738609	<i>LINC02770</i>	CT	C	0.39	$3.57 \times 10^{-8}$	-0.007

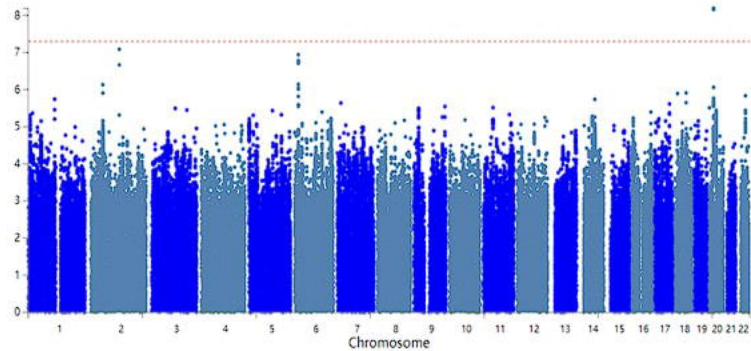
Chr: chromosome



Primary GWAS

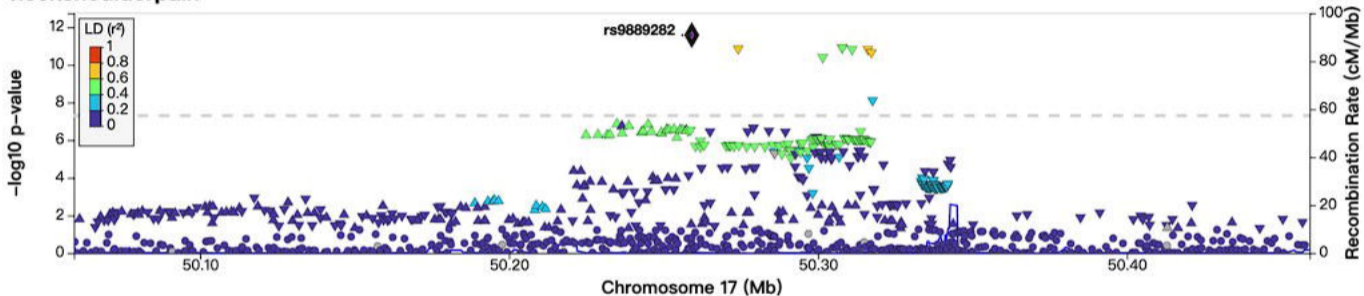


Female-specific GWAS



Male-specific GWAS

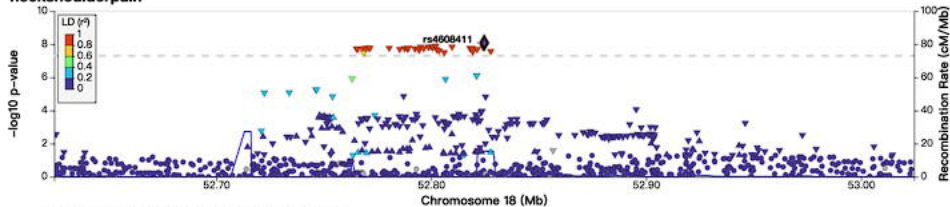
# neckshoulderpain



## GWAS Catalog hits for neckshoulderpain



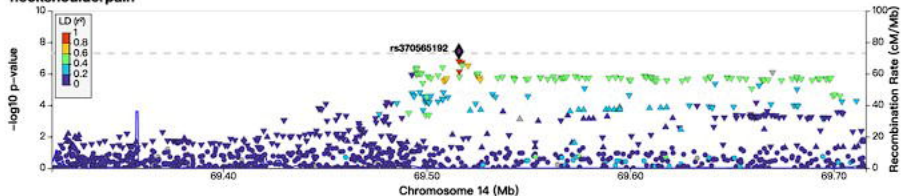
### neckshoulderpain



### GWAS Catalog hits for neckshoulderpain

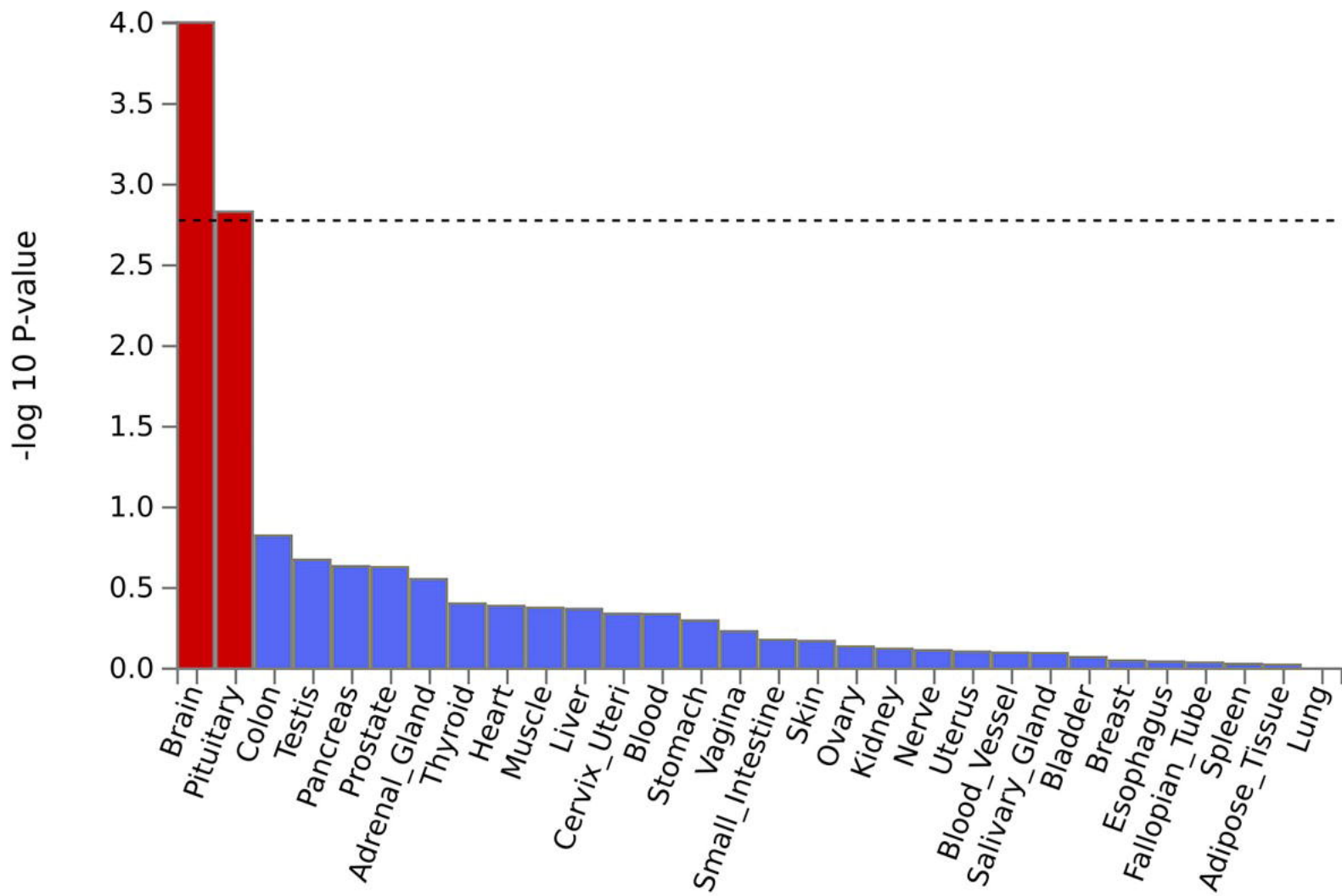


### neckshoulderpain



### GWAS Catalog hits for neckshoulderpain







-log<sub>10</sub> P-value

