Category: Original Article

Running head: GWAS on 'neck or shoulder pain'

A genome-wide association study finds genetic variants associated with neck or shoulder pain in UK Biobank

Authors: Weihua Meng[1*], Brian W Chan[1*], Cameron Harris[1], Maxim B Freidin[2], Harry L Hebert[1], Mark J Adams[3], Archie Campbell[4], Caroline Hayward[5], Hua Zheng[6], Xianwei Zhang[6], Lesley A Colvin[1], Tim G Hales[7], Colin NA Palmer[1], Frances MK Williams[2], Andrew McIntosh[3], Blair H Smith[1].

[1] Division of Population Health and Genomics, Medical Research Institute, Ninewells Hospital and School of Medicine, University of Dundee, Dundee, UK, DD2 4BF

[2] Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences, King's College London, London, UK, SE1 7EH

[3] Division of Psychiatry, Edinburgh Medical School, University of Edinburgh, Edinburgh, UK, EH10 5HF

[4] Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK, EH4 2XU

[5] Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK, EH4 2XU

[6] Department of Anaesthesiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[7] Institute for Academic Anaesthesia, Division of Systems Medicine, School of

Medicine, Ninewells Hospital, University of Dundee, Dundee, UK

* These two authors contributed to this paper equally.

Corresponding Author: Dr Weihua Meng, Ph.D.

Address: Division of Population Health and Genomics, School of Medicine, University of Dundee, Dundee, UK, DD2 4BF.

Tel: +44 1382383419; Fax: +44 1382383802. E-mail: w.meng@dundee.ac.uk

Funding Sources: This study was mainly funded by the Wellcome Trust Strategic Award "Stratifying Resilience and Depression Longitudinally" (STRADL) with reference number 104036/Z/14/Z and by the GCRF academic exchange visits to China funded by the University of Dundee.

Generation Scotland: Scottish Family Health Studies (GS:SFHS) received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006).

TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

Conflicts of Interest: None declared.

Significance: This is the first genome-wide association study on neck or shoulder pain. We have identified 3 genetic loci (an intergenic region in chromosome 17, the *FOXP2* gene in chromosome 7, and the *LINC01572* gene in chromosome 16) that are associated with neck or shoulder pain using the UK Biobank cohort, among

which the *FOXP2* gene and the *LINC01572* gene were weakly replicated by the Generation Scotland: Scottish Family Health Study ($P < 0.05$). The SNP heritability was 0.11, indicating neck or shoulder pain is a heritable trait. The tissue expression analysis suggested that neck or shoulder pain was related to multiple brain tissues, indicating the involvement of neuron function. The results will inform further research in the characterisation of the mechanisms of neck or shoulder pain.

ABSTRACT

Background: Common types of musculoskeletal conditions include pain in the neck and shoulder areas. This study seeks to identify the genetic variants associated with neck or shoulder pain based on a genome-wide association approach using 203,309 subjects from the UK Biobank cohort and look for replication evidence from the Generation Scotland: Scottish Family Health Study (GS:SFHS) and TwinsUK.

Methods: Cases in the UK Biobank were determined by a question which asked the participants if they had experienced pain in the neck or shoulder in the previous month influencing daily activities. Controls were the UK Biobank participants who reported no pain anywhere in the last month. A genome-wide association study was performed adjusting for age, sex, BMI and 9 population principal components. Significant and independent genetic variants were then sent to GS:SFHS and TwinsUK for replication.

Results: We identified 3 genetic loci that were associated with neck or shoulder pain in the UK Biobank samples. The most significant locus was in an intergenic region in chromosome 17, rs12453010, having $P = 1.66 \times 10^{-11}$. The second most significant locus was located in the *FOXP2* gene in chromosome 7 with $P = 2.38 \times 10^{-10}$ for rs34291892. The third locus was located in the *LINC01572* gene in chromosome 16

with $P = 4.50 \times 10^{-8}$ for rs62053992. In the replication stage, among 4 significant and independent genetic variants, rs2049604 in the *FOXP2* gene and rs62053992 in the *LINC01572* gene were weakly replicated in GS:SFHS ($P = 0.0240$ and $P = 0.0202$, respectively). None of the single nucleotide polymorphisms (SNPs) were replicated in the TwinsUK cohort ($P > 0.05$).

Conclusions: We have identified 3 loci associated with neck or shoulder pain in the UK Biobank cohort, two of which were weakly supported in a replication cohort. Further evidence is needed to confirm their roles in neck or shoulder pain.

Key words: neck pain, shoulder pain, GWAS, genetics, FOXP2, UK Biobank

INTRODUCTION

Musculoskeletal pain in the neck and shoulder areas is a major health problem for adults of working age as well as for elderly populations (1). Neck and shoulder pain are prevalent forms of self-reported musculoskeletal pain (2). The aetiologies of neck and shoulder pain may be complicated since both regional lesions and systemic disorders outside the cervicobrachial area may cause pain at that location (3,4). In addition, lesions in the neck can lead to pain in the shoulder and vice versa (5). Many people also have difficulty in describing and differentiating pain in these areas accurately. For these reasons, neck or shoulder pain is often discussed as a single entity (6).

Epidemiological studies have suggested that the prevalence of neck pain is 6-22% and 7-27% for shoulder pain (7–9). The Global Burden of Disease Study 2010 found that of the 291 conditions studied, neck or shoulder pain as a single entity ranked 21st in overall burden on society, and 4th in terms of overall disability (8). The updated Global Burden of Disease Study 2016 also indicated that neck pain was a

top five cause of years lived with disability (YLD) in high-income, and high-middle-income countries.(10) Risk factors associated with neck or shoulder pain conform to the biopsychosocial model; specifically, they include older age, being female, high body mass index (BMI), previous injury, strenuous occupation and diabetes mellitus (3,11–14). Although mechanical exposure is associated with increased risk of pain in the neck and shoulder, this explains only part of these complaints (15). Because of the biopsychosocial factors involved, treating neck or shoulder pain successfully is a challenge. In a study of neck, shoulder and arm pain, only 25% of the patients made a complete recovery after 6 months (16). Estimated rates of remission 1 year after neck or shoulder pain onset were between 33-55% (17–20).

Genetic studies have identified genes associated with neck or shoulder pain. Twin studies have shown that there is a genetic role in neck pain (14), though in keeping with many traits the genetic component becomes smaller with age (21,22). Nonetheless, in adolescents as much as 68% of variance in neck pain liability could be attributed to genetic factors (23). So far there has been no genome-wide association study (GWAS) published on neck or shoulder pain.

This study seeks to identify the genetic variants associated with neck or shoulder pain based on a GWAS approach in a cohort of 203,309 subjects from the UK Biobank cohort and to test significant results for replication in the Generation Scotland: Scottish Family Health Study (GS:SFHS) and TwinsUK. Similar approaches have been used to examine back pain, knee pain, and headaches in the UK Biobank cohort (24–26).

METHODS

Participants and the genetic information of cohorts

Discovery cohort - The UK Biobank (https://www.ukbiobank.ac.uk/) is a project facilitating research into health and disease and involves over 500,000 participants aged between 40-69 years old at recruitment. Participants completed a detailed questionnaire which examined lifestyle, demographic factors, and clinical history. Participants also underwent clinical measures and baseline body measurements such as height and weight. Biological samples including urine, saliva, and blood were also provided. The National Research Ethics Service granted ethical approval to the UK Biobank (reference 11/NW/0382). The genetic information of 500,000 participants was released to approved researchers in March 2018. The corresponding author of this paper was granted access to the genetic information under UK Biobank application number 4844. Detailed quality control information pertaining to these genotypes was described by Bycroft et al (27).

Replication cohort 1 - Generation Scotland: Scottish Family Health Study (GS:SFHS) is a multi-institutional, family-based cohort involving over 20,000 volunteer participants aged between 18-98 years old at recruitment, who provided blood samples from which DNA was extracted. Similarly, participants also completed questionnaires to provide detailed phenotypic and sociodemographic information. Clinical and biochemical measurements were also collected for the purpose of research. Permission was obtained for linkage of research data to routine health data in the form of electronic health records (28,29). Ethical approval for GS:SFHS was obtained from the Tayside Committee on Medical Research Ethics (on behalf of the National Health Service) with reference Number 05/S1401/89. The genetic information relating 20,000 participants was released to the corresponding author of

this paper in March 2018 for pain-related research. Detailed quality control information pertaining to these genotypes was described by Hall et al (30).

Replication cohort 2 - The TwinsUK cohort is a UK nationwide registry of volunteer same sex twins. It has recruited 14,274 registered twins aged between 16 and 98 years. Collection of data and biological materials commenced in 1992 and is ongoing. During study participation, participants regularly complete health and lifestyle questionnaires and visit collaborating clinics and hospitals for clinical evaluation. Ethical approval was provided by the Research Ethics Committee at Guy's and St. Thomas' NHS Foundation Trust. TwinsUK has the genetic information relating to 6,921 participants. Detailed quality control information pertaining to the genetic information was described by Moayyeri et al (31).

Phenotypic definitions on neck or shoulder pain

Discovery cohort - UK Biobank: Participants were offered a pain-related questionnaire, which included the question: 'in the last month have you experienced any of the following that interfered with your usual activities?'. The options were: 1. Headache; 2. Facial pain; 3. Neck or shoulder pain; 4. Back pain; 5. Stomach or abdominal pain; 6. Hip pain; 7. Knee pain; 8. Pain all over the body; 9. None of the above; 10. Prefer not to say. Participants could select more than one option. (UK Biobank Questionnaire field ID: 6159).

In this study, cases were defined as participants who reported having activity limiting pain in the neck or shoulder in the past month (option 3), regardless of whether they reported pain in other regions. The controls were defined as participants who chose the 'None of the above' option.

Replication cohort 1 – GS:SFHS: Participants were first asked ''Are you currently troubled by pain or discomfort, either all the time or on and off?''. If yes was selected,

then the participants were asked ''Have you had this pain or discomfort for more than 3 months?''. If yes was selected once again, then they were asked ''Where is this pain or discomfort?'' with options of 'Back pain', 'Neck or shoulder pain', 'Headache, facial or dental pain', 'Stomach ache or abdominal pain', 'Pain in your arms, hands, hips, legs or feet', 'Chest pain', and 'Other pain'. If a participant selected the 'Neck or shoulder pain', then he/she was defined as a case. All other subjects were defined as controls.

Replication cohort 2: TwinsUK: participants were asked 'In the past three months, have you had pain in your neck or shoulders?' Those who answered 'Yes' were defined as cases. Those who answered 'No' were defined as controls. Those with missing answers were not included in the study.

Statistical Analysis

Discovery cohort – UK Biobank: BGENIE (https://jmarchini.org/bgenie/) was used as the main GWAS software. Single nucleotide polymorphisms (SNPs) with imputation INFO scores < 0.1, minor allele frequency (MAF) < 0.5% were removed, as well as SNPs that failed Hardy-Weinberg tests $P < 1.0 \times 10^{-6}$.

BGENIE was used to perform association studies using linear association tests, adjusting for age, sex, BMI, 9 population principle components, genotyping arrays, and assessment centres. Chi-square testing was used to compare gender difference between cases and controls. T-tests were used to compare age and BMI between case and control groups using IBM SPSS 22 (IBM Corporation, New York). As is standard in GWAS, SNP associations were considered significant when $P < 5.0 \times 10^{-8}$. Genome-wide Complex Trait Analysis (GCTA) software was used to calculate SNP-based or narrow-sense heritability (32). Significant and independent SNPs from the GWAS results of the discovery cohort were then sent to replication cohorts for

replication. These significant and independent SNPs were defined by FUMA with r^2 (linkage disequilibrium score) < 0.6 with any other significant SNPs.

Replication cohort 1 – GS:SFHS: GCTA fastGWA1.92.4 was the main software used for replication. (https://cnsgenomics.com/software/gcta/#fastGWA). SNPs with INFO scores < 0.3, or MAF < 1% were removed, as well as SNPs that failed Hardy-Weinberg tests $P < 1.0 \times 10^{-6}$. FastGWA was used to perform association studies using a mixed-effects linear model adjusting for age, sex, BMI, and 9 population principle components. Relatedness was adjusted for via a genetic kinship matrix.

Replication cohort 2 – TwinsUK: GEMMA v 0.98.1 was used for replication (https://github.com/genetics-statistics/GEMMA). A mixed-effects linear model adjusting for age, sex, BMI, and relatedness via a genetic kinship matrix was used.

Meta-analysis of the significant and independent SNPs combining the UK Biobank, GS:SFHS and TwinsUK was performed using GWAMA 2.2.2 (https://genomics.ut.ee/en/tools/gwama).

Post-GWAS analysis: This study used FUMA as a main annotation tool for viewing and annotating GWAS results (33). It applied SNP functional annotations and generated a corresponding GWAS Manhattan plot.

MAGMA v1.06 (integrated in FUMA) was used to perform gene-based association analysis and gene-set analysis, both of which were generated from GWAS summary statistics (34). For gene-based association analysis, all SNPs located in protein coding genes are mapped to one of 19,123 protein coding genes. The default SNP-wise model (mean) was applied. We tested the joint association of all SNPs in the gene with the phenotype by aggregating the SNP summary statistics to the level of whole genes. In gene-set analysis, individual genes were aggregated to groups of genes sharing certain biological, functional or other characteristics. This aims to

elucidate the involvement of specific biological pathways or cellular functions in the genetic aetiology of a phenotype. GTEx (also integrated in FUMA, https://www.gtexportal.org/home/) provided the results of tissue expression analysis. Genetic correlation analysis was also performed to identify genetic correlation between neck or shoulder pain and 234 complex traits based on the online tool LD hub v1.9.0 (http://ldsc.broadinstitute.org/ldhub/). Any *P* value less than 2.1 x 10^{-4} (0.05/234) was considered statistically significant by Bonferroni adjusted testing.

RESULTS

GWAS Results

In the UK Biobank, 775,252 responses to all options were received for the specific pain question. Of the 501,708 participants in the study, 123,061 participants reported having experienced activity limiting pain in the neck or shoulder in the previous month. 213,408 participants chose the 'None of the above' option which meant they did not have activity limiting pain anywhere in the previous month. To create a homogeneous dataset, we first removed samples according to their ancestry information. In addition, those who were related to one or more others in the cohort (a cut-off value of 0.044 in the generation of the genetic relationship matrix) and those who failed quality control were also removed. The final number of those included in the case group after the above exclusions was 53,994 (28,093 males, 25,901 females). 149,312 (71,480 males, 77,832 females) individuals were included in the control group. After SNP quality control, there were 9,304,965 SNPs available for GWAS analysis. Clinical characteristics of the case and control groups were compiled (Table 1). Age, sex, and BMI were all found to be significantly different ($P < 0.001$) between cases and controls. Three genetic loci including 4 significant and

independent SNPs reached a GWAS significance of $P < 5\times10^{-8}$ (Figure 1, Table 2). The most significant locus was located in an intergenic region in chromosome 17. The SNP from this location of highest significance was rs12453010 ($P = 1.66 \times 10^{-10}$). The second locus was found in the *FOXP2* gene located in chromosome 7, and the most significant SNP from this locus was rs34291892 ($P = 2.38 \times 10^{-10}$). The third locus was the *LINC01572* gene located in chromosome 16, and the most significant SNP in this locus was rs62053992 ($P = 4.50 \times 10^{-10}$). The SNP heritability (liability scale) from GCTA was 0.11±0.017.

These 4 significant and independent SNPs were tested for replication in GS:SFHS and TwinsUK. Among 20,032 subjects in the GS:SFHS subjects, 19,632 had complete relevant information. The whole-genome fastGWA results did not find any SNPs with GWAS significance associated with neck or shoulder pain. (Supplementary Figure S1). Among the 4 SNPs from the discovery cohort, rs2049604 in the *FOXP2* gene and rs62053992 in the *LINC01572* gene were replicated weakly ($P = 0.0240$ and 0.0202, respectively) (Table 2). Of the 6,921 individuals in TwinsUK with genetic information, 3,982 of them had valid relevant information on phenotypes and covariates. None of the 4 SNPs were replicated in the TwinsUK cohort ($P > 0.05$).

Meta-analysis of the 4 significant and independent hits, combining UK Biobank, GS:SFHS and Twins UK found that the significance of their associations was increased (Table 2).

FUMA Analysis

In gene-based association analysis by MAGMA, 26 genes were found to be associated with neck or shoulder pain, all of which are represented in Supplementary Table S1. The most significant gene was *FOXP2* ($P = 1.62 \times 10^{-11}$), which is located

in chromosome 7.

In gene-set analysis conducted by MAGMA, 10,651 gene-sets were analysed using the default competitive test model. None of these gene-sets met genome-wide significance ($P < 4.7 \times 10^{-6}$ (0.05/10651)).

Tissue expression analysis was conducted by GTEx, and the relationship between tissue specific gene expression and genetic associations was tested by using the average gene expression in each tissue type as a covariate. Two analyses were carried out, one investigating 30 general tissue types (Figure 2) and the other looking at 53 specific tissue types (Figure 3). Tissue expression analysis in 30 tissue types found expression in brain tissue to be the most significant ($P = 9.53 \times 10^{-5}$). Only expression in brain and pituitary tissue reached significant values of $P < 1.67 \times 10^{-3}$ (0.05/30). Tissue expression analysis of 53 specific tissue types by GTEx found expression in the nucleus accumbens of the basal ganglia to be the most significant ($P = 3.55 \times 10^{-5}$). In addition, the top 6 significant associations were all from brain tissues.

The genetic correlation analysis using the LD hub (v1.9.0) showed that symptoms of depression had the largest significant and positive genetic correlation with neck or shoulder pain (rg = 0.5522, $P = 3.41 \times 10^{-30}$), followed by insomnia (rg = 0.5377, $P = 1.21 \times 10^{-21}$). It was also found that the age at which they have their first child had the most significant and negative genetic correlations with neck or shoulder pain (rg = -0.4812, $P = 8.95 \times 10^{-37}$), followed by college completion (rg = -0.4706, $P = 7.26 \times 10^{-26}$). All the phenotypes with significant genetic correlation with neck or shoulder pain are shown in Table 3.

DISCUSSION

We have performed a GWAS on neck or shoulder pain using the UK Biobank resource and found 3 loci that have reached genome-wide significance (P <0.05 × 10^{-8}). They are the *FOXP2* gene in chromosome 7, the *LINC01582* gene in chromosome 16 and an intergenic area in chromosome 17. In the replication stage, the *FOXP2* and *LINC01582* loci were weakly supported by the GS:SFHS cohort, but not by TwinsUK.

FOXP2 belongs to the forkhead-box transcription factor family and encodes a 715 amino acid long transcription factor (35). It may have 300-400 transcription targets, and has a forkhead/winged helix binding domain with 2 polyglutamine tracts adjacent to each other due to a mixture of CAG and CAA repeats (36,37). *FOXP2* is a gene shown to be vital in the neural mechanisms underpinning the development of speech and language. A previous study described a family with developmental verbal dyspraxia, where affected individuals had a loss of function mutation in the *FOXP2* gene (38). Affected individuals in the family had difficulty with the selection and sequencing of fine orofacial muscular movements needed to articulate words, as well as deficits in language and grammatical skills. *FOXP2* also plays a role in regulating 'hub' genes *Dlx5* and *Syt4* in animal models, which are important for brain development and function (39-41). The mutations in the *FOXP2* gene were also associated with decreased grey matter in the cerebellum (42).

Notably, *FOXP2* is expressed in several regions of the brain, namely the basal ganglia, locus coeruleus, parabrachial nucleus, and thalamus. All of these regions have been previously implicated in the modulation of pain (43–46), and also concur with the tissue expression analysis (Figure 2, Figure 3) which suggests that the central nervous system modulates neck or shoulder pain. A recent GWAS study suggested that *FOXP2* is a candidate gene for multisite chronic pain (47). Further

studies into the location and function of the transcription targets of *FOXP2* would also provide valuable insight.

The LINC01572 gene in chromosome 16 was also replicated in this study, with rs62053992 having the *P* value at 4.5×10^{-8} in the discovery cohort and at 0.02 in the GS:SFHS. The gene is 384kb long, and was recently suggested to be related with polycystic ovary syndrome (48). However, studies of this gene have been very limited. Further replication evidence should be sought to confirm its role in neck or shoulder pain. Although the locus in chromosome 17 was not replicated, it was the most significantly associated in the discovery cohort with rs12453010 having the lowest *P* value of 1.66×10^{-11}. Analysis of this locus is also difficult as it is a gene desert area. The nearest neighbouring gene is *CA10,* which encodes a protein belonging to the carbonic anhydrase family and is responsible for catalysing the hydration of carbon dioxide. It is also thought to contribute to central nervous system development particularly in the brain (49).

The results of the meta-analysis of the 4 significant and independent SNPs combining 3 cohorts suggested that the loci identified by the UK Biobank cohort were supported by the GS:SFHS and TwinsUK. It is likely that the sample size in the TwinsUK (N= 3,982) was too small to replicate the loci for such a heterogeneous phenotype as neck or shoulder pain. In particular, for the 4 SNPs chosen for replication, the power to achieve a *P* value < 0.05 in the given TwinsUK sample ranged between 11.6% to 15.8% (50).

The SNP heritability of neck or shoulder pain was 0.11. This is similar to that of back pain (0.11), greater than knee pain (0.08), and less than: hip pain (0.12), stomach or abdominal pain (0.14), headache (0.21), facial pain (0.24), and pain all over the body (0.31) (51). Further, the genetic correlation matrix among 8 pain phenotypes in the

UK Biobank showed that the neck or shoulder pain and back pain shared the highest genetic correlation (rg = 0.83) (51).

The genetic correlation analysis results were perhaps to be expected. Like knee pain and back pain, which have been shown to be positively correlated with depression and neuroticism (51), neck or shoulder pain was correlated genetically and positively with some mental health and personality phenotypes. We also identified that neck or shoulder pain was genetically and negatively correlated with the age at which they have their first child, college completion, and years of schooling. This means that those who were older when they had their first child, those with more years of schooling, and those with completed college education were less likely to report neck or shoulder pain. These factors could be related to a number of factors including lifestyle, deprivation levels, and occupation. It is interesting to note that males are more likely to report neck or shoulder pain than females in the UK Biobank population. This is matched with the fact that males are more likely to have strenuous occupations. However, we should note, in general, female sex is a risk factor for neck or shoulder pain.

The primary limitation of this study is that different (albeit similar) case and control definitions were used in the discovery and replication cohorts. This was a consequence of the pre-determined phenotypic information that was present in the relevant cohorts. We defined neck or shoulder pain cases and controls based on the responses by UK Biobank participants to a specific pain question. This question focused on neck or shoulder pain occurrence during the previous month that was sufficient to cause interference with activity. The severity, frequency, and exact location of the neck or shoulder pain were not documented. Hence, our phenotyping should be considered as broadly defined. In the GS:SFHS and Twins UK cohorts,

the disease status of participants was also self-reported, while cases were those who having neck or shoulder pain over the past 3 months, controls were those who could have neck or shoulder pain less than 3 months or have pain in other body sites. While in the UK Biobank, controls were defined as pain free for the past month. Differences between (albeit similar) the case and control definitions could have a negative impact on replication of the results while this impact is hard to evaluate Differences between the case and control definitions could have a negative impact on replication of the results although this impact would be hard to evaluate. There could also be some cases who report neck or shoulder pain as a result of underlying causes such as cancer and osteoarthritis in the neck and shoulder areas. Their impact to the results would be very limited due to the small numbers.

In summary, we have identified 3 loci of genome-wide significance ($P < 5\times10^{-8}$) associated with neck or shoulder pain in the UK Biobank dataset using a GWAS approach. Two of these loci were replicated weakly in the GS:SFHS cohort. Identification of these loci now provides a foundation for future work into understanding genetic roles and aetiology in neck or shoulder pain.

ACKNOWLEDGEMENTS

We would like to thank all participants of the UK Biobank, Generation Scotland and TwinsUK cohorts who have provided necessary genetic and phenotypic information.

This research has been conducted using the UK Biobank Resource under Application Number 4844.

Generation Scotland is grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and

399 laboratory technicians, clerical workers, research scientists, volunteers, managers,
400 receptionists, healthcare assistants, and nurses.
401
402 **Data availability**
403 The GWAS summary statistics of neck or shoulder pain can be accessed through
404 https://figshare.com/articles/fourpainphenotype2/7699583
405
406 **Conflict of interest**
407 The authors declare that they have no conflict of interest.
408
409 **AUTHOR CONTRIBUTIONS**
410 WM organised the project, drafted the paper and contributed to the analysis. BWC
411 and CH contributed to drafting the paper. MBF provided the TwinsUK replication. MA
412 performed the main UK Biobank GWAS analysis. AC and CH contributed to
413 GS:SFHS cohorts. HLH, HZ, XZ, CNAP, LC, TGH and FMKW provided essential
414 comments. AM and BS organised the project and provided comments.

REFERENCES

1. Vogt, M. T., Simonsick, E. M., Harris, T. B., et al. Neck and shoulder pain in 70- to 79-year-old men and women: findings from the Health, Aging and Body Composition Study. *Spine J.*, **3**, 435–441.
2. Picavet, H. S. J. and Schouten, J. S. A. G. (2003) Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC(3)-study. *Pain*, **102**, 167–178.
3. Slater, M., Perruccio, A. V and Badley, E. M. (2011) Musculoskeletal comorbidities in cardiovascular disease, diabetes and respiratory disease: the impact on activity limitations; a representative population-based study. *BMC Public Health*, **11**, 77.
4. Molsted, S., Tribler, J. and Snorgaard, O. (2012) Musculoskeletal pain in patients with type 2 diabetes. *Diabetes Res. Clin. Pract.*, **96**, 135–140.
5. Fernández-de-Las-Peñas, C., Galán-Del-Río, F., Alonso-Blanco, C., et al. (2010) Referred pain from muscle trigger points in the masticatory and neck-shoulder musculature in women with temporomandibular disoders. *J. Pain*, **11**, 1295–1304.
6. Rossi, M., Pasanen, K., Kokko, S., et al. (2016) Low back and neck and shoulder pain in members and non-members of adolescents' sports clubs: the Finnish Health Promoting Sports Club (FHPSC) study. *BMC Musculoskelet. Disord.*, **17**, 263.
7. Wright, A. R., Shi, X. A., Busby-Whitehead, J., et al. (2009) The Prevalence of Neck and Shoulder Symptoms and Associations with Comorbidities and Disability: The Johnston County Osteoarthritis Project. *Myopain*, **23**, 34–44.
8. Hoy, D., March, L., Woolf, A., et al. (2014) The global burden of neck pain:

estimates from the global burden of disease 2010 study. *Ann. Rheum. Dis.*, **73**, 1309–1315.

9. March, L., Smith, E. U. R., Hoy, D. G., et al. (2014) Burden of disability due to musculoskeletal (MSK) disorders. *Best Pract. Res. Clin. Rheumatol.*, **28**, 353–366.

10. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators (2017) Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet (London, England)*, **390**, 1211–1259.

11. Nilsen, T. I. L., Holtermann, A. and Mork, P. J. (2011) Physical exercise, body mass index, and risk of chronic pain in the low back and neck/shoulders: longitudinal data from the Nord-Trondelag Health Study. *Am. J. Epidemiol.*, **174**, 267–273.

12. Webb, R., Brammah, T., Lunt, M., et al. (2003) Prevalence and predictors of intense, chronic, and disabling neck and back pain in the UK general population. *Spine (Phila. Pa. 1976).*, **28**, 1195–1202.

13. Miranda, H., Viikari-Juntura, E., Martikainen, R., et al. (2001) A prospective study of work related factors and physical exercise as predictors of shoulder pain. *Occup. Environ. Med.*, **58**, 528–534.

14. MacGregor, A. J., Andrew, T., Sambrook, P. N., et al. (2004) Structural, psychological, and genetic influences on low back and neck pain: a study of adult female twins. *Arthritis Rheum.*, **51**, 160–167.

15. Ostergren, P.-O., Hanson, B. S., Balogh, I., et al. (2005) Incidence of shoulder and neck pain in a working population: effect modification between mechanical

and psychosocial exposures at work? Results from a one year follow up of the Malmö shoulder and neck study cohort. *J. Epidemiol. Community Health*, **59**, 721–728.

16. Feleus, A., Bierma-Zeinstra, S. M. A., Miedema, H. S., et al. (2007) Prognostic indicators for non-recovery of non-traumatic complaints at arm, neck and shoulder in general practice--6 months follow-up. *Rheumatology (Oxford).*, **46**, 169–176.

17. Bot, S. D. M., van der Waal, J. M., Terwee, C. B., et al. (2005) Incidence and prevalence of complaints of the neck and upper extremity in general practice. *Ann. Rheum. Dis.*, **64**, 118–123.

18. Enthoven, P., Skargren, E. and Oberg, B. (2004) Clinical course in patients seeking primary care for back or neck pain: a prospective 5-year follow-up of outcome and health care consumption with subgroup analysis. *Spine (Phila. Pa. 1976).*, **29**, 2458–2465.

19. Hoving, J. L., de Vet, H. C. W., Twisk, J. W. R., et al. (2004) Prognostic factors for neck pain in general practice. *Pain*, **110**, 639–645.

20. Vos, C. J., Verhagen, A. P., Passchier, J., et al. (2008) Clinical course and prognostic factors in acute neck pain: an inception cohort study in general practice. *Pain Med.*, **9**, 572–580.

21. Fejer, R., Hartvigsen, J. and Kyvik, K. O. (2006) Heritability of neck pain: a population-based study of 33,794 Danish twins. *Rheumatology (Oxford).*, **45**, 589–594.

22. Hartvigsen, J., Petersen, H. C., Pedersen, H. C., et al. (2005) Small effect of genetic factors on neck pain in old age: a study of 2,108 Danish twins 70 years of age and older. *Spine (Phila. Pa. 1976).*, **30**, 206–208.

490 23. Ståhl, M. K., El-Metwally, A. A., Mikkelsson, M. K., et al. (2013) Genetic and
491 environmental influences on non-specific neck pain in early adolescence: a
492 classical twin study. *Eur. J. Pain*, **17**, 791–798.
493 24. Meng, W., Adams, M. J., Palmer, C. N. A., et al. (2019) Genome-wide
494 association study of knee pain identifies associations with GDF5 and
495 COL27A1 in UK Biobank. *Commun. Biol.*, **2**, 321.
496 25. Meng, W., Adams, M. J., Hebert, H. L., et al. (2018) A Genome-Wide
497 Association Study Finds Genetic Associations with Broadly-Defined Headache
498 in UK Biobank (N=223,773). *EBioMedicine*, **28**, 180–186.
499 26. Suri, P., Palmer, M. R., Tsepilov, Y. A., et al. (2018) Genome-wide meta-
500 analysis of 158,000 individuals of European ancestry identifies three loci
501 associated with chronic back pain. *PLoS Genet.*, **14**, e1007601.
502 27. Bycroft, C., Freeman, C., Petkova, D., et al. (2018) The UK Biobank resource
503 with deep phenotyping and genomic data. *Nature*, **562**, 203–209.
504 28. Smith, B. H., Campbell, H., Blackwood, D., et al. (2006) Generation Scotland:
505 the Scottish Family Health Study; a new resource for researching genes and
506 heritability. *BMC Med. Genet.*, **7**, 74.
507 29. Smith, B. H., Campbell, A., Linksted, P., et al. (2013) Cohort profile: Generation
508 scotland: Scottish family health study (GS: SFHS). The study, its participants
509 and their potential for genetic research on health and illness. *Int. J. Epidemiol.*
510 30. Hall, L. S., Adams, M. J., Arnau-Soler, A., et al. (2018) Genome-wide meta-
511 analyses of stratified depression in Generation Scotland and UK Biobank.
512 *Transl. Psychiatry*, **8**, 9.
513 31. Moayyeri, A., Hammond, C. J., Hart, D. J., et al. (2013) The UK Adult Twin
514 Registry (TwinsUK Resource). *Twin Res. Hum. Genet.*, **16**, 144–149.

515 32. Yang, J., Lee, S. H., Goddard, M. E., et al. (2011) GCTA: a tool for genome-
516 wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
517 33. Watanabe, K., Taskesen, E., van Bochoven, A., et al. (2017) Functional
518 mapping and annotation of genetic associations with FUMA. *Nat. Commun.*, **8**,
519 1826.
520 34. de Leeuw, C. A., Mooij, J. M., Heskes, T., et al. (2015) MAGMA: generalized
521 gene-set analysis of GWAS data. *PLoS Comput. Biol.*, **11**, e1004219.
522 35. Lai, C. S. L., Fisher, S. E., Hurst, J. A., et al. (2001) A forkhead-domain gene is
523 mutated in a severe speech and language disorder. *Nature*, **413**, 519–523.
524 36. Spiteri, E., Konopka, G., Coppola, G., et al. (2007) Identification of the
525 transcriptional targets of FOXP2, a gene linked to speech and language, in
526 developing human brain. *Am. J. Hum. Genet.*, **81**, 1144–1157.
527 37. Vernes, S. C., Spiteri, E., Nicod, J., et al. (2007) High-throughput analysis of
528 promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in
529 speech and language disorders. *Am. J. Hum. Genet.*, **81**, 1232–1250.
530 38. Vargha-Khadem, F., Watkins, K., Alcock, K., et al. (1995) Praxic and nonverbal
531 cognitive deficits in a large family with a genetically transmitted speech and
532 language disorder. *Proc. Natl. Acad. Sci. U. S. A.*, **92**, 930–933.
533 39. Konopka, G., Bomar, J. M., Winden, K., et al. (2009) Human-specific
534 transcriptional regulation of CNS development genes by FOXP2. *Nature*, **462**,
535 213–217.
536 40. Acampora, D., Merlo, G. R., Paleari, L., et al. (1999) Craniofacial, vestibular
537 and bone defects in mice lacking the Distal-less-related gene Dlx5.
538 *Development*, **126**, 3795–3809.
539 41. Yoshihara, M., Adolfsen, B., Galle, K. T., et al. (2005) Retrograde signaling by

Syt 4 induces presynaptic release and synapse-specific growth. *Science*, **310**, 858–863.

42. Belton, E., Salmond, C. H., Watkins, K. E., et al. (2003) Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia. *Hum. Brain Mapp.*, **18**, 194–200.

43. Yen, C.-T. and Lu, P.-L. (2013) Thalamus and pain. *Acta Anaesthesiol. Taiwan*, **51**, 73–80.

44. Brightwell, J. J. and Taylor, B. K. (2009) Noradrenergic neurons in the locus coeruleus contribute to neuropathic pain. *Neuroscience*, **160**, 174–185.

45. Benarroch, E. E. (2006) Pain-autonomic interactions. *Neurol. Sci.*, **27 Suppl 2**, S130-5133.

46. Chudler, E. H. and Dong, W. K. (1995) The role of the basal ganglia in nociception and pain. *Pain*, **60**, 3–38.

47. Johnston, K. J. A., Adams, M. J., Nicholl, B. I., et al. (2019) Genome-wide association study of multisite chronic pain in UK Biobank. *PLoS Genet.*, **15**, e1008164.

48. Zhao, J., Xu, J., Wang, W., et al. (2018) Long non-coding RNA LINC-01572:28 inhibits granulosa cell growth via a decrease in p27 (Kip1) degradation in patients with polycystic ovary syndrome. *EBioMedicine*, **36**, 526–538.

49. 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. *Biochim. Biophys. Acta*, **1518**, 311–316.

50. Visscher, P. M., Wray, N. R., Zhang, Q., et al. (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.*, **101**, 5–22.

51. Meng, W., Adams, M. J., Reel, P., et al. (2019) Genetic correlations between

565 pain phenotypes and depression and neuroticism. *Eur. J. Hum. Genet.*

566

LEGENDS

Figure 1: The Manhattan plot of the GWAS on neck or shoulder pain using the UK Biobank cohort (N=203,309). The dashed red line indicates the cut-off *P* value of 5×10^{-8}.

Figure 2: 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.

Figure 3: 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.

Figure 1

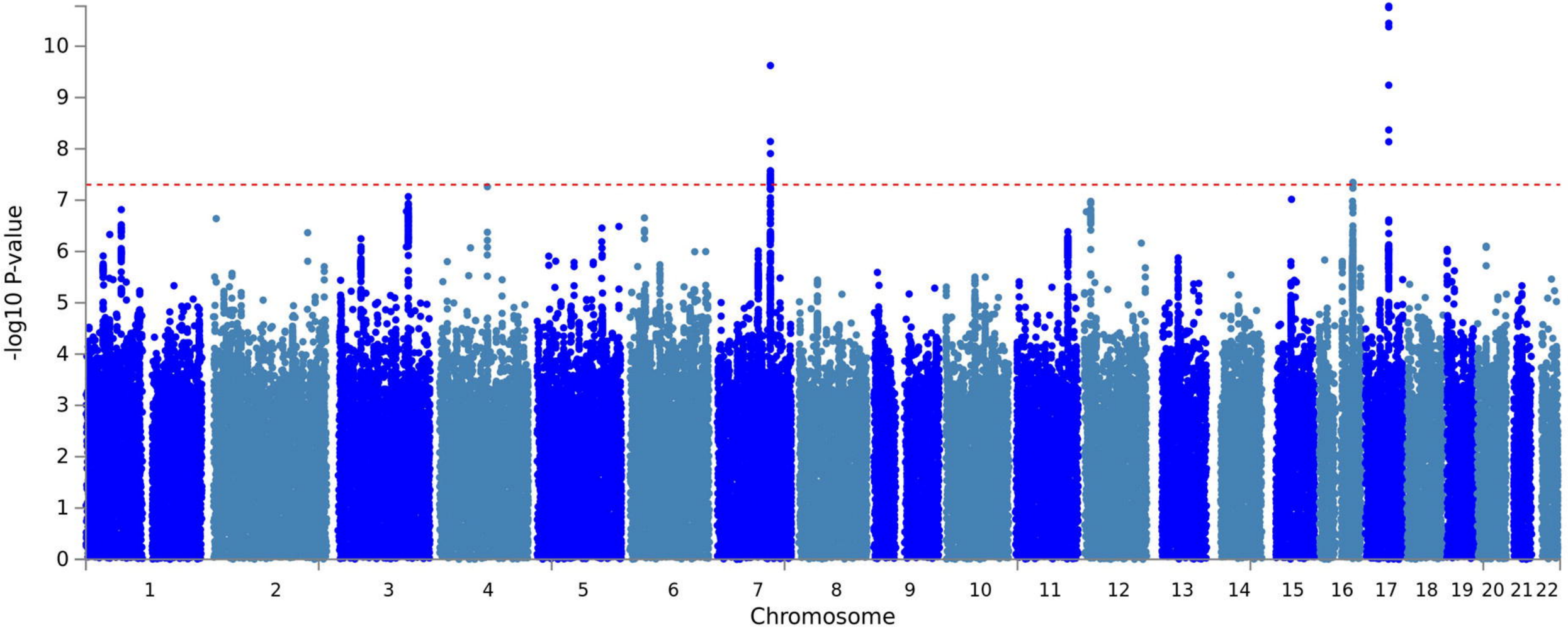

Figure 2

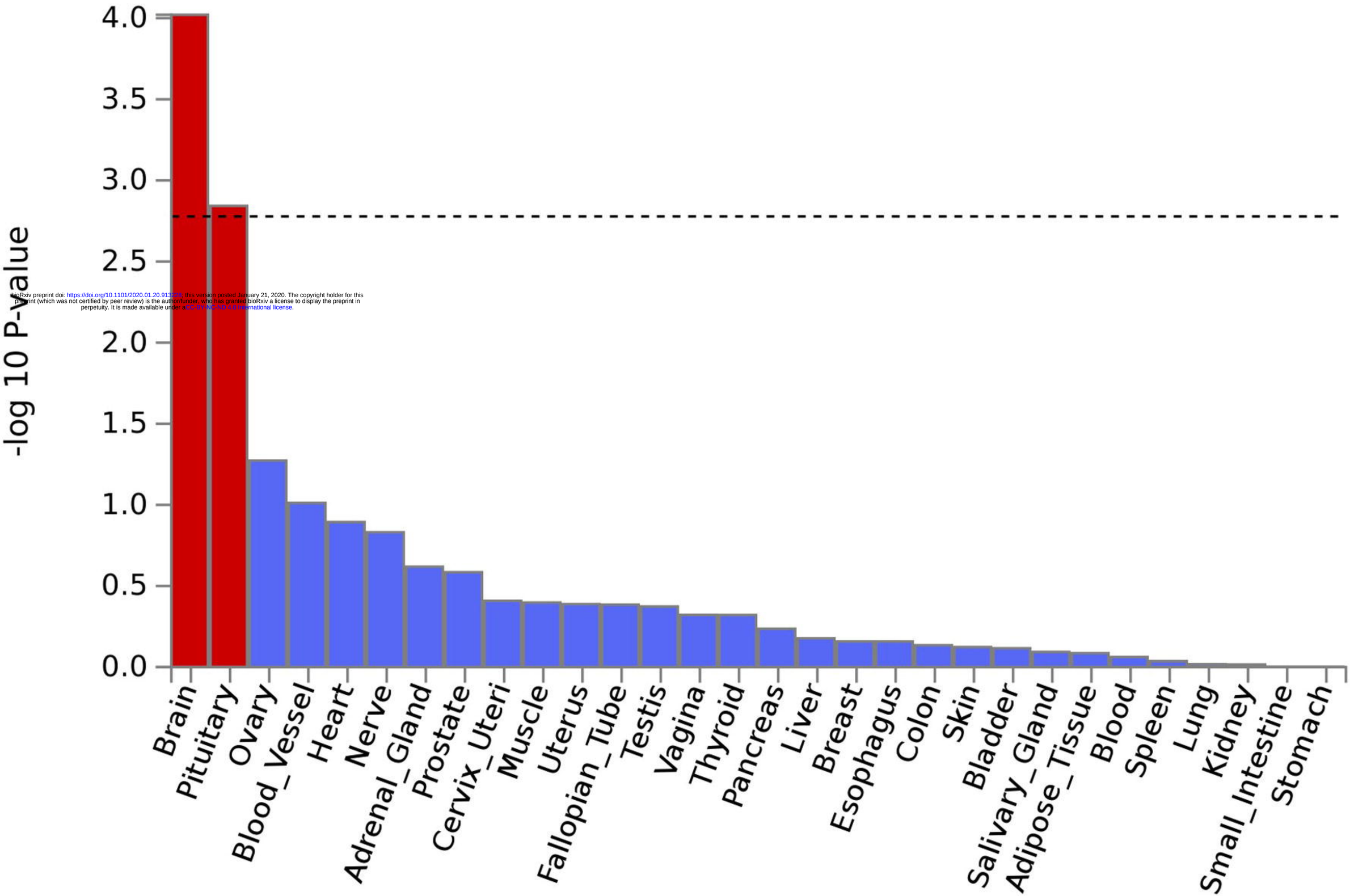

Figure 3

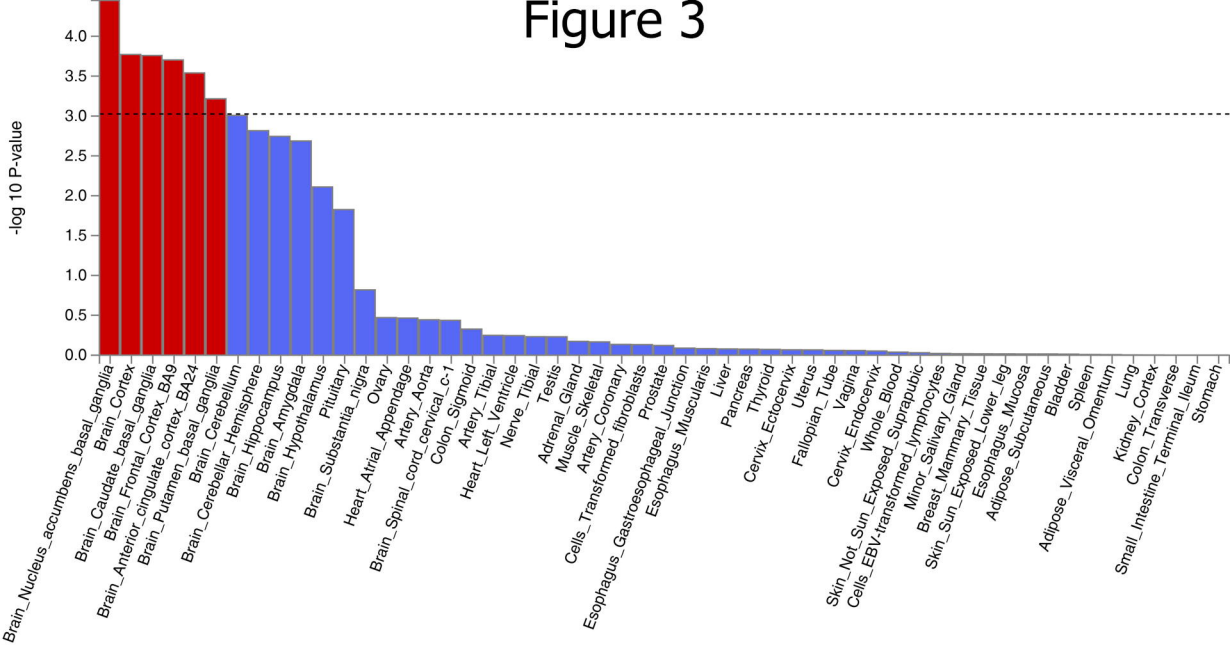
-log 10 P-value
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Brain_Nucleus_accumbens_basal_ganglia
Brain_Cortex
Brain_Caudate_basal_ganglia
Brain_Frontal_Cortex_BA9
Brain_Anterior_cingulate_cortex_BA24
Brain_Putamen_basal_ganglia
Brain_Cerebellum
Brain_Cerebellar_Hemisphere
Brain_Hippocampus
Brain_Amygdala
Brain_Hypothalamus
Pituitary
Brain_Substantia_nigra
Ovary
Heart_Atrial_Appendage
Artery_Aorta
Brain_Spinal_cord_cervical_c-1
Colon_Sigmoid
Artery_Tibial
Heart_Left_Ventricle
Nerve_Tibial
Testis
Adrenal_Gland
Muscle_Skeletal
Artery_Coronary
Cells_Transformed_fibroblasts
Prostate
Esophagus_Gastroesophageal_Junction
Esophagus_Muscularis
Liver
Pancreas
Thyroid
Cervix_Ectocervix
Uterus
Fallopian_Tube
Vagina
Cervix_Endocervix
Whole_Blood
Skin_Not_Sun_Exposed_Suprapubic
Cells_EBV-transformed_lymphocytes
Minor_Salivary_Gland
Breast_Mammary_Tissue
Skin_Sun_Exposed_Lower_leg
Esophagus_Mucosa
Adipose_Subcutaneous
Bladder
Spleen
Adipose_Visceral_Omentum
Lung
Kidney_Cortex
Colon_Transverse
Small_Intestine_Terminal_Ileum
Stomach

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

	UK Biobank		
Covariates	Cases	Controls	*P*
Sex (male:female)	28093 : 25901	71,480 : 77,832	<0.001
Age (years)	57.7 (7.82)	56.9 (7.97)	<0.001
BMI (kg/m^2)	27.8 (4.85)	26.7 (4.30)	<0.001

BMI: body mass index

A chi-square 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).

Table 2. The summary statistics of the 4 significant and independent SNPs in the UK Biobank and the replication and meta-analysis results using the GS:SFHS and TwinsUK cohorts

			UK Biobank Discovery stage				GS:SFHS replication			TwinsUK replication			Meta-analysis		
SNP	Chr	Position	MAF	*P*	beta	se	*P*	beta	se	*P*	beta	se	*P*	effects	I^2
rs2049604	7	113990352	0.3648 (T)	3.26×10^{-8}	-0.0080	0.0015	**0.0240**	-0.0088	0.0039	0.13	-0.0188	0.0122	3.19×10^{-9}	---	0
rs34291892	7	114058731	0.3772 (C)	2.38×10^{-10}	-0.0092	0.0014	0.1068	-0.0064	0.0039	0.11	-0.0196	0.0124	7.07×10^{-12}	---	0
rs62053992	16	72389872	0.1819 (G)	4.50×10^{-8}	-0.0098	0.0018	**0.0202**	-0.0109	0.0047	0.94	0.0012	0.0158	4.36×10^{-9}	--+	0
rs12453010	17	50316131	0.3926 (T)	1.66×10^{-11}	0.0095	0.0014	0.1331	0.0058	0.0038	0.10	0.0199	0.0121	2.20×10^{-12}	+++	0

rs34291892 was replaced by rs4727799 (r2=0.8736) as it does not exist in the GS:SFHS and TwinsUK cohorts.

UK Biobank N=203,309; GS:SFHS N= 19,632; TwinsUK N= 3,982

Chr: chromosome

MAF: minor allele frequency

Significant *P* values in the replication ($P < 0.05$) were in bold.

Table 3. The significant genetic correlation results by the LD hub between neck or shoulder pain with other phenotypes

Trait 1	Trait 2	rg	*P*
Neck or shoulder pain	Depressive symptoms	0.5522	3.41×10^{-30}
Neck or shoulder pain	Insomnia	0.5377	1.21×10^{-21}
Neck or shoulder pain	Neuroticism	0.4417	2.00×10^{-29}
Neck or shoulder pain	Major depressive disorder	0.3968	5.75×10^{-08}
Neck or shoulder pain	Ever vs never smoked	0.3082	4.14×10^{-08}
Neck or shoulder pain	Number of children ever born	0.2616	2.03×10^{-07}
Neck or shoulder pain	Coronary artery disease	0.1965	1.66×10^{-07}
Neck or shoulder pain	Waist-to-hip ratio	0.1547	1.31×10^{-05}
Neck or shoulder pain	Sleep duration	-0.2468	9.46×10^{-07}
Neck or shoulder pain	Subjective well being	-0.2626	1.14×10^{-06}
Neck or shoulder pain	Intelligence	-0.3229	2.23×10^{-17}
Neck or shoulder pain	Former vs Current smoker	-0.3600	5.87×10^{-06}
Neck or shoulder pain	Years of schooling 2016	-0.4373	5.18×10^{-54}
Neck or shoulder pain	College completion	-0.4706	7.26×10^{-26}
Neck or shoulder pain	Age of first birth	-0.4812	8.95×10^{-37}

rg: genetic correlation

The cut-off *P* value is 0.0002 (0.05/234).