MASTER OF SCIENCE

A Genome-Wide Association Study on Widespread Pain using the UK Biobank Cohort

Fang, Jimin

Award date:
2021

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.

• You may not further distribute the material or use it for any profit-making activity or commercial gain

• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
A Genome-Wide Association Study on Widespread Pain using the UK Biobank Cohort

Author: Jimin Fang
Degree: Master of Science (MSc) by Research
Contents

Acknowledgement ......................................................................................................................... 6
Declaration .................................................................................................................................. 6
Abbreviations ............................................................................................................................... 7
Abstract ........................................................................................................................................ 8

1. Introduction .............................................................................................................................. 9

1.1. Definition & classification of fibromyalgia (FM) & chronic widespread pain (CWP) ................................................................. 9

1.2. Prevalence of FM & CWP .................................................................................................... 12

1.3. Aetiology & pathophysiology of FM & CWP ................................................................. 14

1.3.1. Related laboratory evaluations ..................................................................................... 14

1.3.2. Neuroimaging investigations ........................................................................................ 15

1.3.3. Concept of central sensitisation .................................................................................... 17

1.3.4. Animal models ................................................................................................................ 18

1.3.5. Genetic factors in FM & CWP ....................................................................................... 18

1.3.6. Personality factors in FM & CWP ................................................................................ 21

1.3.7. Psychosocial stressors related to FM & CWP ............................................................... 23

1.4. Management of FM & CWP ............................................................................................... 24

1.4.1. Medication ...................................................................................................................... 25

1.4.2. Non-pharmacological treatment .................................................................................... 27

1.5. Introduction to the UK Biobank ......................................................................................... 29

1.5.1. The UK Biobank Axiom genotyping array .................................................................... 30

1.6. Objective of the study .......................................................................................................... 31

2. Methodology ............................................................................................................................ 32

2.1. Basics of GWAS ................................................................................................................ 32

2.1.1. GWAS based on unrelated individuals .......................................................................... 34

2.1.2. Multiple test adjustment in GWA studies ..................................................................... 34

2.1.3. Limitations in GWAS .................................................................................................... 36

2.2. Participants and their genetic information ......................................................................... 37
List of Tables

Table 1-1 Different criteria of FM & CWP (pain all over the body) (1990-2016). 11
Table 1-2 The London FM Epidemiology Study Screening Questionnaire (LFESSQ) used in FM prevalence surveys. 12
Table 1-3 Summary of management of FM. 28
Table 3-1 Genders, ages and BMIs of cases and controls of CWP. 43
Table 3-2 The summary statistics of the three significant and independent SNPs. 45
Table 3-3 The significant gene sets with Bonferroni p value < 0.05. 45
Table 4-2 FAF1 protein reactions (adopted from Craig W. Menges). 56
List of Figures

Figure 1-1 The locations of the nine paired tender points that constitute the 1990 ACR criteria for FM ................................................................. 10

Figure 1-2 Worldwide distribution of FM prevalence in general population from 2005–2014 (Maraques AP et al.) ............................................................. 13

Figure 1-3 The first fMRI study in FM ..................................................... 16

Figure 1-4 The Big Five personality traits (OCEAN model) ...................... 22

Figure 1-5 Summary of the UK Biobank resource and genotyping array content (Bycroft C et al.) ................................................................. 30

Figure 2-1 General classification of GWA studies .................................... 33

Figure 2-2 Common methods of multiple testing corrections ................... 35

Figure 2-4 Flowsheet of the methodology ........................................... 39

Figure 2-5 Basic functions of SNP2GENE of FUMA ......................... 40

Figure 3-1 Original & Adjusted Statistics of Cases & Controls in the UK Biobank ................................................................. 43

Figure 3-2 The Manhattan plot of the GWAS on CWP using the UK Biobank ...... 44

Figure 3-3 The Q-Q plot of GWAS summary statistics ........................ 46

Figure 3-4 The Q-Q plot of the gene-based test computed by MAGMA .... 46

Figure 3-5 The Manhattan plot of the gene-based test as computed by MAGMA based on the input GWAS summary statistics ......................... 49

Figure 3-6 Tissue expression results on 30 specific tissue types by GTEx in the FUMA ................................................................. 50

Figure 3-7 Tissue expression results on 53 specific tissue types by GTEx in the FUMA ................................................................. 51

Figure 4-1 Locations of LINC01078 and TBC1D4 gene in chromosome 13 .... 53

Figure 4-2 Location of FAF1 gene in chromosome 1 .............................. 53
Acknowledgement

I am appreciative of the generous support from my supervisors, Dr Weihua Meng, Prof Lesley Colvin, and Mrs Jennifer Watson. Without their help, the project would not be complete.

Declaration

I declare that the content of this project report is my own work and has not previously been submitted for any other assessment. The report is written in my own words and conforms to the University of Dundee’s Policy on plagiarism and academic dishonesty. Unless otherwise indicated, I have consulted all the references cited in this report.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACR</td>
<td>American College of Rheumatology</td>
</tr>
<tr>
<td>CWP</td>
<td>chronic widespread pain</td>
</tr>
<tr>
<td>eQTL</td>
<td>expression quantitative trait locus</td>
</tr>
<tr>
<td>FAF1</td>
<td>FAS-associated factor 1</td>
</tr>
<tr>
<td>FM</td>
<td>fibromyalgia</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional magnetic resonance imaging</td>
</tr>
<tr>
<td>FMS</td>
<td>fibromyalgia syndrome</td>
</tr>
<tr>
<td>FUMA</td>
<td>functional mapping and annotation of genome-wide association studies</td>
</tr>
<tr>
<td>GSA</td>
<td>gene-set analysis</td>
</tr>
<tr>
<td>GWAS</td>
<td>genome-wide association study</td>
</tr>
<tr>
<td>HRQoL</td>
<td>health-related quality of life</td>
</tr>
<tr>
<td>MAGMA</td>
<td>multi-marker analysis of GenoMic annotation</td>
</tr>
<tr>
<td>MHC</td>
<td>major histocompatibility complex</td>
</tr>
<tr>
<td>WPI</td>
<td>widespread pain index</td>
</tr>
</tbody>
</table>
Abstract

Background: Pain affecting multiple sites on the body is one of the main symptoms of fibromyalgia (FM) and it can be a complementary phenotype to other site-specific pains. The aim of this study was to explore the genetic variants associated with widespread pain in a cohort of UK Biobank participants who had completed a pain phenotyping questionnaire.

Method: Cases and controls were determined by a question answered by UK Biobank participants. The question contained 10 selections to identify different pain sites around the body. Those selecting ‘pain all over the body’ (9,551) formed the case group, and those selecting ‘none of the above’ (232,467) formed the control group. Cases were removed based on ancestry information, ethnicity, and other possible confounders. 5,670 cases and 149,312 controls met inclusion criteria. A genome-wide association study (GWAS) on pain all over the body was performed adjusting with age, sex and body mass index (BMI). The GWAS summary statistics were then annotated by the online annotation software FUMA.

Results: Based on the significance level of \( p < 0.05 \times 10^{-8} \), nine GWAS-tagged candidate SNPs were found. These SNPs were located in 3 loci. In the first locus, the SNP rs17387024 found at the FAF1 gene in chromosome 1 was most significant with a \( p\)-value of \( 3.72 \times 10^{-8} \). In the second locus, the most significant SNP was rs550883786 at the intergenic region in chromosome 4 with a \( p\)-value of \( 3.65 \times 10^{-8} \). In the third locus, located at LINC01078 in chromosome 13, the most significant SNP was rs148500993, with a \( p\)-value of \( 3.65 \times 10^{-8} \). According to related biochemical researches, FAF1 gene encodes FAF1 (FAS-associated factor 1) which plays an important role in apoptotic mechanisms by extrinsic and intrinsic pathways, and furthermore FAF1 can intervene the sensing of pain indirectly.

Conclusions: The results suggest that pain all over the body is related to altered brain function, and the steps to realise the mechanisms of this relationship lie with further characterisation of the three genomic loci identified as well as medical imaging results.
1. Introduction
Chronic widespread pain (CWP), also known as widespread chronic pain (WCP), is a chief symptom of fibromyalgia (FM). CWP is a particular type of chronic pain in which the pathophysiological causality is not fully understood and cannot be simply identified as either neuropathic or musculoskeletal pain. (1) Self-report questionnaires may refer to CWP as ‘pain all over the body’

1.1. Definition & classification of fibromyalgia (FM) & chronic widespread pain (CWP)
Since the initial categorization of CWP by the American College of Rheumatology (ACR) 1990 criteria as a subset of FM, coded as WP1990, the definition of CWP has undergone a number of changes and updates. (2) At first, the authors that created the criteria sought a pain variable (an early version of widespread pain index, WPI) to identify patients with FM. Patients then underwent follow-up examinations related to tender points, by which the so-called ‘widespread pain’ was defined as ‘pain above and below the waist, on both sides of the body and in the axial region (see Figure 1-1^2), and with no exclusions as to the source of the pain’ (WP1990) (2, 3). At that time, WP1990 was supplementary to the FM diagnosis rather than a separate condition. However, physicians in the following years had tended to use such a definition rather widely, to diagnose patients, resulting in an increased incidence. By 2015 there were more than 1,500 identified citations to CWP, and CWP had also become a discrete entity and subject in many reviews. This resulted in appeals for a more precise definition and more prudent application of the diagnosis. (4) Though it was frequently used in clinical practice, WP1990 was imprecise at times: the number of tender points did not have a strict limit (5, 6), and patients with limited somatic pain (i.e., as few as three tender points) could also be diagnosed with CWP, which diverged from the original

1 see https://www.ukbiobank.ac.uk/wp-content/uploads/2019/08/Pain-questionnaire-for-website_Copyright.pdf
2 From https://commons.wikimedia.org/wiki/File:Tender_points_FM.gif
definitions. The inherent deficiency of WP1990 led to alternative definitions. (7, 8) In 2016, the definition was updated, and the name of CWP was changed into ‘generalised pain’ in the report in order not to be confused with the previous criteria. The presence of at least four or five body regions was emphasised, which tried to tally with FM diagnosis. (7) Experts argue that the definition of CWP should not only function as a criterion for FM and could be used outside the context of FM. Meanwhile, there are considerations that CWP should be a single diagnostic category and become a hypernym of FM in ICD-11. (1, 9, 10) Nevertheless, it is now still accepted that CWP is a multifactorial pain condition characterised by prolonged pain which lasts for three months or more in multiple regions of the body. According to the current criteria, FM diagnosis also requires the symptom of CWP as well as significant psychophysiological distress in the form of anxiety, anger, frustration, depression, insomnia, and social isolation that can be quantified by polysymptomatic distress scale (PSD). (11-15) Table 1-1 lists the changes in definition and criteria for FM and CWP respectively since 1990 by American College of Rheumatology. (3)

Figure 1-1 The locations of the nine paired tender points that constitute the 1990 ACR criteria for FM
Table 1-1 Different criteria of fibromyalgia (FM) & chronic widespread pain (CWP) (pain all over the body) (1990-2016)

<table>
<thead>
<tr>
<th>Code name</th>
<th>Definition &amp; Details</th>
<th>Code name</th>
<th>Definition &amp; Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>WP1990</td>
<td>Pain is considered widespread when all the following are present: pain in the left side of the body, pain in the right side of the body, pain above the waist, and pain below the waist. In addition, axial skeletal pain (cervical spine or anterior chest or thoracic spine or low back) must be present a.</td>
<td>FM2010</td>
<td>A patient satisfies the diagnostic criteria for FM if the following conditions are met — (1) widespread pain index (WPI) ≥ 7 and symptom severity score (SSS) ≥ 5 or WPI 3-6 and SSS ≥ 9; (2) symptoms have been present at a similar level for at least 3 months.</td>
</tr>
<tr>
<td>WP2016</td>
<td>Developed as a widespread pain criterion for FM diagnosis, it is satisfied by the presence of pain in four or five body regions. To avoid confusion with the WP1990 criteria, it was called ‘generalized pain’ in the 2016 report. A proposed criterion modified from WP2016 to function as a standalone categorical measure of CWP. It is satisfied by pain in four or five body regions (in effect the WP2016 criterion) in the presence of at least seven painful sites. a. A mandatory component of the ACR 1990 FM criteria. b. Categorical widespread pain assessment is not contained.</td>
<td>FM2011</td>
<td>A modification of the ACR 2010 FM criteria that allows for self-report b.</td>
</tr>
<tr>
<td>WP2019</td>
<td></td>
<td>FM2016</td>
<td>A modification of FM2011, FM diagnosis requires (A) WPI ≥ 7 and SSS ≥ 5 OR a WPI 4–6 and an SSS ≥ 9, (B) presence of WP2016, and (C) symptoms of at least 3 months duration.</td>
</tr>
</tbody>
</table>
1.2. Prevalence of FM & CWP

FM is generally considered as a common syndrome that appears in clinical practice and highly impacts the health-related quality of life (HRQoL). There have been clinical studies undertaken in various countries, and the prevalence of FM in the general population is approximately 3%, ranging from 0.4% to 9.3% worldwide. (16-18)

In the early surveys, 1990 American College of Rheumatology criteria were almost the only source of definition of FM, and the criteria were applied widely. The criteria of pain and fatigue listed in Table 1-1 is adapted from the 1990 ACR criteria. According to a research from 1998, the overall prevalence of FM in the adult population in the United States was estimated at 2.0% (95% confidence interval [CI]: 1.4 to 2.7); prevalence was lower in men (0.5%) than in women (3.4%) and increased with age. (19) According to a Canadian community survey, FM affects 3.3% (95% CI: 3.2 to 3.4) of adults in London, Ontario, being more prevalent in women than in men. (8) Similar results of prevalence have been reported in Western European countries, including Germany (3.0%, 95% CI: 1.6 to 4.4), Spain (2.4%, 95% CI: 1.5 to 3.2), Italy (2.2%; 95% CI: 1.4 to 3.2), and Sweden (2.5%; 95% CI not provided) (20-23). Conversely, the prevalence of FM was found to be as low as 0.8% (95% CI not provided) in Finland (24) and 0.7% (95% CI: 0.3 to 1.3) in Denmark (25). Another survey was performed in five European countries (France, Germany, Italy, Portugal, and Spain) with the use of LFESSQ-4 (the London FM Epidemiology Study Screening Questionnaire, meeting the 4-pain criteria alone) and LFESSQ-6 (meeting the 4-pain and 2-fatigue criteria).

Table 1-2 The London FM Epidemiology Study Screening Questionnaire (LFESSQ) used in FM prevalence surveys

<table>
<thead>
<tr>
<th>Pain criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the past 3 months:</td>
</tr>
<tr>
<td>1. Have you had pain in muscles, bones, or joints, lasting at least 1 week?</td>
</tr>
<tr>
<td>2. Have you had pain in your shoulders, arms, or hands? On which side? Right, left, or both?</td>
</tr>
</tbody>
</table>
3. Have you had pain in your legs or feet? On which side? Right, left, or both?
4. Have you had pain in your neck, chest or back?

Meeting the pain criteria requires ‘yes’ responses to all
4 pain items, and either (A) both a right- and left-side positive response, or (B) a both sides positive response.

<table>
<thead>
<tr>
<th>Fatigue criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Over the past 3 months, do you often feel tired or fatigued?</td>
</tr>
<tr>
<td>6. Does tiredness or fatigue significantly limit your activities?</td>
</tr>
</tbody>
</table>

Screening positive for chronic, debilitating fatigue requires a ‘yes’ response to both fatigue items.

The estimated overall prevalence of FM was 4.7% (95% CI: 4.0 to 5.3) and 2.9% (95% CI: 2.4 to 3.4), respectively, in the general population. The prevalence of FM was age- and sex-related and varied among countries. (26) Figure 1-2 depicts a bibliographical survey carried out from 2005 to 2014, including 39 studies. The 2010 American College of Rheumatology criteria had not been widely used yet. (27)

A precise definition of FM and CWP has yet to be agreed, which has potentially resulted in inaccurate or incomplete diagnosis. In addition, practitioners from different professional specialties (urology, gastroenterology, etc.) tend to make different diagnoses (chronic prostatitis, irritable bowel syndrome, etc.) which are referred to as chronic overlapping pain conditions or functional disorders, though the symptoms of those patients may share the same pathology. (28)

Figure 1-2 Worldwide distribution of FM prevalence in general population from 2005–2014
One method to quantify the epidemiological information of FM is the London FM Epidemiology Study Screening Questionnaire (LFESSQ). It is mainly used for FM screening in general population surveys of noninstitutionalised adults. (29) The details of LFESSQ are shown in Table 1-2.

1.3. **Aetiology & pathophysiology of FM & CWP**

The definitions proposed by ACR help to interpret and categorise FM & CWP, but the cause of these diseases still remains unclear. (30) Generally, it is hypothesised that the aetiology of FM is a complex interaction of genetic (31) neurological (28, 32, 33), immunological (34) and environmental (35) factors.

1.3.1. **Related laboratory evaluations**

There is no specific laboratory test for FM. The erythrocyte sedimentation rate (ESR) and other inflammatory indices are expected to be normal. The anti-nuclear antibodies (ANA) and rheumatoid factors are typically negative and do not warrant specific FM screening tests. There might be incidental cases of positive results which take place due to the increasing ages of patients. Other diseases related to CWP, such as endocrine disorders and Lyme disease, should be considered during clinical practice and excluded from the diagnosis of FM, while relevant serological tests are not routinely required. (36)
1.3.2. Neuroimaging investigations

The altered central neural processing among patients with FM experience was firstly depicted by neuroimaging techniques. (37) The objective information from neuroimaging helped to convince the sceptics that pain in FM is ‘real’, and the therapeutic rapport between patients and physicians can be ensured. Most common functional imaging techniques used in FM are single-photon emission computed tomography, functional magnetic resonance, positron emission tomography and proton magnetic resonance spectroscopy.

Single-photon emission computed tomography (SPECT) is a nuclear medicine tomographic imaging technique using γ rays. SPECT is the first functional neuroimaging technique used in FM studies. Guedj et al. found hyper-perfusion of a sensitive radioligand (radioactive biochemical substance, a ligand that is radio-labelled) in the somatosensory cortex, and hypo-perfusion in the cingulate, amygdala, medial frontal, and parahippocampal gyrus, and cerebellum. (38)

Functional magnetic resonance (fMRI) supports central augmentation of sensory input in FM. (39, 40) The results of the first fMRI study in FM are shown in Figure1-3 below. (28) In fMRI studies, the brain regions that most strongly encode for stimulus intensity are the posterior insula and the secondary somatosensory cortices, and these are the brain regions where neuronal activation will be most accentuated in individuals with diffuse hyperalgesia or allodynia which encompasses symptoms of FM. (41-43) Petzke et al. used fMRI to evaluate the central pain processing of patients with FM after the milnacipran treatment, revealing a trend of increasing pressure-pain tolerance after the treatment, though it was not statistically significant. (44)

Positron emission tomography (PET) aims to measure brain metabolism and the distribution of exogenous radiolabelled chemical agents throughout the brain. Harris et al used PET to show that the μ-opioid receptors in patients with
FM were less available. (45) Wood et al compared patients with FM to healthy controls. (46) After noxious intramuscular administration of hypertonic saline infusion, PET showed that only the control group responded to the painful stimuli by releasing dopamine in the basal ganglia, which illustrated a lack of response in the dopaminergic reward system to stress-induced analgesia. (47) It provides objective evidence that patients with FM have an abnormal dopamine response to pain. (48)

Figure 1-3 The first fMRI study in FM

Proton magnetic resonance spectroscopy (1H-MRS) uses natural hydrogen to determine the structure of molecules, illustrating how pharmacological and non-

3 Individuals with FM (in red triangle) were given a low intensity stimulus (shown in top left panel) and this led to moderate pain (a 0–20 Gracely scale was used to rate pain intensity). Their fMRI BOLD responses were compared to controls given approximately the same intensity stimulus (blue box) or a higher intensity stimulus that was necessary to cause the same amount of pain (green circle). No significant neuronal activation was shown from this low intensity stimulus in the controls, but there was in patients with FM, and these areas of neuronal activation overlapped significantly with the brain activation pattern of the controls given nearly twice as much pressure, which was what was required to cause comparable amounts of pain.
pharmacological therapies contribute to analgesia among patients. (49) Harris et al. investigated 10 patients with FM by $^1$H-MRS and fMRI before and after a non-drug intervention to reduce pain. Validated pain scores were used before each imaging session. The changes in glutamate levels measured in the insula reflected changes in pain domains. (50)

In conclusion, there have been various medical imaging studies carried out in FM patients, with identified neuropathological pathways supporting the theory of central sensitisation. By improving our understanding of the molecular processes involved in FM and somatisation disorder, there is the potential to develop novel targeted treatments to improve the prognosis of FM patients.

1.3.3. Concept of central sensitisation

Hyperalgesia (sensitivity to pain increases abnormally) and/or allodynia (ordinarily nonpainful stimuli elicit pain) can be presented amongst most patients with FM. Central sensitisation, an emerging biopsychosocial concept, acknowledge the diffuse and widespread nature of the pain as the key feature of FM, and other frequent comorbidities such as fatigue, dyssomnia, mood disorder (regarded as ‘central sensitivity syndrome’) as the abnormal amplification of pain in the central nervous system that possibly results in the development of FM. Due to central sensitisation, the electrophysiological discharge in sensory processing is prolonged, and various forms of stimulation (e.g. smell, noise, chemical exposure, etc.) are also exaggerated. It has been discovered that several areas in the central nervous system that are responsible for reducing ascending pain transmission within the spinal cord (e.g., brain stem, cortico-reticular system, locus coeruleus and hypothalamus) through the activity of inhibitory neurotransmitters including serotonin, norepinephrine, enkephalins, γ-aminobutyric acid and adenosine. (51) In some other chronic pain studies, an imbalance between descending inhibitory and facilitatory systems is suggested to contribute to central sensitisation. (52, 53)
1.3.4. Animal models

Animal models have been developed to emulate the neurobiological process of CWP i.e., hyperalgesia and/or allodynia, and models with repeated irritations to muscles have become most common and best characterised. Sluka et al. attempted to induce a non-inflammatory pain model by repeated intramuscular injections of acidic saline (pH = 4.0). The model produced widespread hyperalgesia of skin, muscle, and viscera while observable tissue damage and inflammation were avoided. (54) The model was replicated by other researchers in following years (55, 56), and an association was found between the induction of the model and a 50-60% incidence of anxiety-like and depression-like behaviours. (57)

Diffuse noxious inhibitory controls (DNIC), also known as conditioned pain modulation (CPM) is an endogenous pain modulatory pathway which has often been described as ‘pain inhibits pain’, which is believed to have relations to the analgesic effects. (58) DNIC has been widely studied in animals. (59) Different nociceptive conditioning stimuli (electrical stimulation, hot water, acupuncture, or pinch) have been used to induce DNIC in animals. (60-62) By blocking the opioid receptors in the medullary reticularis nucleus dorsalis, de Resende et al found that DNIC increased pain thresholds in uninjured animals and was altered in pain conditions. (63)

1.3.5. Genetic factors in FM & CWP

Significant efforts have been made to uncover the genetic contribution to FM & CWP, with potential roles for of a series of candidate genes in the pathogenesis of FM. Researchers have attempted to explore the genetic essence of FM by assessing the possibility of linkage to the human leukocyte antigens (HLA). Some older studies were conducted before the definition and criteria of FM by ACR (i.e., WP1990) was devised. Some found statistical significance between HLA region and FM. (64, 65) However, such relationships were not confirmed by other studies.
The changing diagnostic criteria and different baselines of population might lead to inconsistent results. Later, the direction of the research was directed to the metabolism of catecholamines, such as receptors of serotonin (68-70), dopamine (71, 72) as well as norepinephrine. The significant roles of the chemicals above in both pain-inhibiting and pain-initiating pathways were likely to correlate with FM, and their functions were grounded on the expression of certain genes and the types of cells. However, the results could not fully explain the aetiology of FM. For example, Frank et al. investigated HTR3A and HTR3B for sequence variations in patients with FM syndrome with a hypothesis that dysfunction of serotonergic neurotransmission could lead to FM syndrome, but the data were not statistically significant and the aetiology of FM was yet to confirm. Lee et al. conducted a candidate gene study with a sample of 60,367 total participants from 237 clinics in the US, 2713 of which had been diagnosed with FM. The study revealed the associations between demography and FM diagnosis in a diverse population as well as some overrepresentations of minor alleles which make up the Catechol-O-methyltransferase (COMT) haplotypes among the population with FM. However, the COMT haplotypes related to pain sensitivity were not directly relevant to FM diagnosis because there were no statistically significant associations of COMT haplotypes or diplotypes with FM diagnosis in the FM group compared to the non-FM group.

Other candidate genes of certain nociceptive neurotransmitters like substance P (SP) have also been investigated. In early studies, SP was explicitly shown to be elevated in level in the cerebral spinal fluid (CSF) among FM patients. SP was also postulated to be involved in chronic stress response and pain signalling because it was one of the old neuropeptides in a view of phylogenetics. Ablin et al. conducted a candidate gene study in the year of 2009 of the 1354 G>C polymorphism in the TACR1 (substance P receptor) but the result did not reach the criteria of statistical significance. Nevertheless, serological research conducted in 2016 showed that substance P along with corticotropin-
releasing hormone (CRH) and SP-structurally-related hemokinin-1 (HK-1) had elevated levels in FMS patients compared with healthy controls, suggesting that these molecules could be released centrally and may cause focal inflammation leading to activation of mast cells, while peripheral inflammation may still have central effects. When mast cells were stimulated, they secreted inflammatory cytokines (IL-6 and TNF) that contribute to the symptoms of FMS. The authors suggested the treatment directed at preventing the secretion or antagonising these elevated neuroimmune markers, both centrally and peripherally. (80)

Genome-wide association study (GWAS) is a revolutionary technique that has dramatically increased the capacity to study the genetic basis of complex traits, including the fields in FM and chronic pain. However, it is still difficult for GWA studies to identify loci that can fully clarify the heritable component. Rare variants possibly result in complex traits and identified loci can only explain a small portion of them. (81) Family-based designs of genetic studies were proved to be useful to detect genetic variants with a large effect size, which tend to be rare in the population, but they are more frequent in cohorts of multi-case families. Using this strategy, Arnold et al. genotyped members of 116 families from the FM family study and a model-free genome-wide linkage analysis of FM syndrome with 341 microsatellite markers was performed. (82) The study recognised chromosome 17p11.2–q11.2 region as a possible major locus for FM syndrome. In particular, the region coincides with two potential candidate genes of FM, viz TRPV2 (transient receptor and the vanilloid channel 2 gene) and SLC6A4 (serotonin transporter gene), while other candidate genes, including COMT, are irrelevant to the region. (83)

Peters et al. have performed a GWAS of CWP, not specifically FM. (84) In the research 1,308 female patients suffering from CWP were tested, along with 5,791 controls. The study also replicated the effects of the genetic variants with evidence for relevance to 1,480 CWP cases and 7,989 controls. The results
demonstrated a genetic variant on chromosome 5p15.2 associated with CWP, which is located upstream at the gene CCT5 (chaperonin-containing-TCP1-complex-5 gene) and downstream at FAM173B (also ATP synthase c subunit lysine N-methyltransferase, ATPSCKMT). The authors remarked that both these genes could be promising targets in the regulation of pain. There were other previous researches that could support the findings in the GWAS. A research in a human pedigree showed that the mutation in CCT5 gene could cause hereditary sensory neuropathy, as known as 'autosomal recessive mutilating sensory neuropathy with spastic paraplegia'. (85) Pain was one of the observed symptoms in the research. Kubota H et al expounded the biochemical functions of CCT5 that CCT5 assisted in folding and assembly protein in the brain as a subunit of the chaperonin containing t-complex polypeptide 1 (TCP-1). (86) Malecki J et al. used bioinformatics analyses and biochemical assays to determine the functions of FAM173B, and uncovered that FAM173B contained an atypical, non-cleavable mitochondrial targeting sequence responsible for its localisation to mitochondria. They also identified FAM173B as the long-sought KMT (mitochondrial lysine-specific methyltransferase) responsible for methylation of ATPSc (ATP synthase c-subunit), a key protein in cellular ATP production, and demonstrated its influential function of ATPSc methylation. (87)

Overall, the results were scarcely concordant with each other due to the limited sample size and sparse replication. Generally, candidate gene studies are often biased by previous hypotheses and non-genetic proofs. There is still a lack in the pathophysiological knowledge about CWP, which increases the chances of inconsistent results and reduces the chances of success.

1.3.6. Personality factors in FM & CWP

The personality of patients with FM & CWP is one of the subjects that arouse much controversy. Some researches revealed specific traits, (88, 89) but others argued that those traits were not differentiated from the normal population. (90)
It is also considered that certain emotional regulation deficits instead of particular traits characterised FM. Historically, patients with FM have been perceived to be close to Cluster B personality disorders (PDs, e.g. dramatic, impulsive and emotional), while emotional dysregulation or self-regulation deficit could help to interpret the similarity. Meanwhile, FM patients show a high level of psychopathology and childhood trauma, as in borderline personality disorder (BPD).

**Figure 1-4 The Big Five personality traits (OCEAN model)**

Based on the Big Five personality traits (also known as OCEAN model) (see Figure 1-4), the Big Five Inventory is one the powerful tools to detect the personality factors in social sciences. It also helps to uncover the personality factors that intervene in the emotional regulation and modulation of pain. Bucourt et al. used the Big Five Inventory to analyse the personality factors of 163 women with FM, rheumatoid arthritis, spondyloarthritis and Sjögren’s syndrome. Patients with FM had higher scores on agreeableness, neuroticism, and openness than those with other rheumatic diseases, which highlighted the specificity of personality in FM.

A systematic review conducted by C Conversano et al. found that many studies underline high levels of alexithymia and type D personality in FM patients, but these results do not differ from those of healthy controls when depression is controlled.
1.3.7. Psychosocial stressors related to FM & CWP

As a main feature of CWP, pain all over the body is of great genetic susceptibility, and yet its development may also be involved with environmental factors. Early-life events, such as physical trauma and psychosocial stressors have been found to interfere in the genetic expression and further contribute to the occurrence. (94, 95)

Studies have been investigating the earliest stage childhood that can form long-term psychological and behavioural alterations. Physical trauma in early life has been proved as a significant factor. Early and childhood experiences, when well quantified and evaluated, have been related to enduring changes in nociceptive circuitry and increased pain sensitivity in older organism. (96) For instance, premature delivery (97) and sexual abuse (98) may contribute to change the pain threshold in adulthood and trigger the development of CWP. (96) Genetic searches also help to elucidate the mechanism. A gene named MAOA (monoamine oxidase A gene) has been shown to alter the effect of childhood maltreatment at risk of antisocial behaviours. As a result, there could be a rise in the impairment of HPA (hypothalamic-pituitary-adrenal) axis, and the response to stress could be less efficient subsequently and the patients are much more sensitive to pain and/or fatigue.

Repeated physical stressors among adults have also been demonstrated to be involved in the development of CWP. Mechanical injury, resulted from activities like heavy lifting, repetitive motions, or squatting for extended periods of time in the workplace, plays an important role in the onset of CWP. (99)

Meanwhile, researches pointed out that psychosocial stressors seem to be the environmental triggers of CWP, and emotional trauma is another strong predictor of the disease in addition to physical trauma. (100, 101) A study in 2005 offered
a snapshot of FM among a self-selected population by an Internet survey, and the result was heuristic that emotional precipitants might trigger FM and/or low back pain. (102) Emotional abuse, post-traumatic stress disorder (PTSD) and other negative environmental conditions are basic characteristics of childhood maltreatment which has been recently discovered to affect FM in another self-reported research. Interestingly, these patients with FM had significantly higher concomitant levels of depression and anxiety. (103) Chang et al. conducted a nationwide longitudinal study and demonstrated a bidirectional temporal association between FM and depression, implying their shared pathophysiology. (104) To support the connection, the morphometric study of altered white and grey matter including medial orbitofrontal cortex and cerebellum was conducted among patients with FM, and the volume of grey matter was found to be related to the severity of hyperalgesia and depression. In addition, the number of fibres in white matter between specific submodule regions was also related to measures of reduced pain sensitivity and clinical pain interference. (105)

Stressful life events in patients with FM demonstrated the transcultural robustness in relation to retrospective self-reports of childhood maltreatment and lifelong traumatic experiences. (106) There is little tangible evidence of physiological processes mediating the connection between the occurrence of FM/CWP and experiencing stress. (107) The hypothalamic-pituitary-adrenal (HPA) axis may influence this relationship: the increased pain levels have been found to be related to the failure of HPA axis and decreased levels of hypothalamic corticotrophin-releasing hormone in the cerebrospinal fluid (CSF). (108)

1.4. Management of FM & CWP

FM is regarded as one of the medically unexplained physical symptoms (MUPS), whose cause remains contested. Thus, a multimodal approach is recommended to be established. (109) Meanwhile, there is a lack of high-quality randomised controlled trials (RCTs) demonstrating convincing efficacy.
Treatment for FM typically encompasses symptom management, alleviating pain, and other comorbidities. Development in the comprehension of the pathophysiology of the disorder have led to improvement in treatment, which may include prescription medication, behavioural intervention, and exercise. In practice, multidisciplinary care is utilised, for instance, comprised of medication, aerobic exercise, cognitive behavioural therapy, etc. have been shown to be effective in alleviating pain and other FM-related symptoms. (110, 111)

1.4.1. Medication
Antidepressants are used to treat major depressive orders, some anxiety disorders, some chronic pain conditions, and to assist in managing addictions. (112) It is reported that antidepressants are ‘associated with improvements in pain, depression, fatigue, sleep disturbances, and health-related quality of life in people with FM syndrome’. (113) The chief goal of antidepressants should be symptom reduction and their effects should also be evaluated against side effects. Placebo-controlled clinical trials have shown that a small number of subjects can benefit significantly from the serotonin–norepinephrine reuptake inhibitors (SNRIs), for example, duloxetine and milnacipran as well as the tricyclic antidepressants (TCAs, e.g., amitriptyline). (114-117) However, many people also encountered adverse events during or after the medication. Evidence is of limited quality that the benefits and harms of selective serotonin reuptake inhibitors (SSRIs) appear to be similar. (118) SSRIs may be used to treat depression in people diagnosed with FM. (119) Tentative evidence suggests that monoamine oxidase inhibitors (MAOIs) such as pirlindole and moclobemide are moderately effective for reducing pain. Very low-quality evidence implies that pirlindole is more effective at treating pain compared with moclobemide. Side effects of MAOIs may include nausea and vomiting. (120)

Anticonvulsants (also commonly known as antiepileptic drugs or as antiseizure
drugs) are used in the treatment of epileptic seizures. Anticonvulsants are also used in the treatment of bipolar disorder (121) and neuropathic pain (122). Some anticonvulsants such as pregabalin and gabapentin demonstrate a small benefit over placebo in alleviating pain and sleep problems in FM. (123) Derry et al. found that pregabalin was able to produce a major reduction in pain intensity over 12 to 26 weeks with tolerable adverse events for a small proportion of people (about 10% more than placebo) with moderate or severe pain due to FM, which resembles the results of other researches in SNRIs (124, 125). Meanwhile, more participants experienced an adverse event with pregabalin (65% with pregabalin vs. 49% with placebo). (126) In another study, provisional evidence shows that gabapentin may be useful for pain in about 18% of people with FM, (127) but it is still difficult to anticipate who will benefit. Cases of side effects such as vertigo also occur among people who take gabapentin. (128) There are also increasing concerns around the abuse potential of gabapentinoids and association with drug deaths. (129)

It is controversial to use opioids to treat FM; though opioids are used worldwide, the U.S. Food and Drug Administration (FDA) does not generally approve their use in this condition, (130) and opioids scarcely take effect. (131) Authorities have concern for the misuse of opioids in FM treatment especially when there are alternative medications to control the pain. They suggest that people using opioids should be placed under surveillance by healthcare providers to prevent side effects and possible unwanted drug behaviours. (132) Long term use of opioids is not currently recommended for the treatment of chronic pain.4

In addition to the aforementioned medications, there are other types of drugs applied to FM treatment. A systematic review in 2014 revealed a possible connection between the disturbances of the proportion of growth hormone (GH) to

4 See https://www.iasp-pain.org/Advocacy/Content.aspx?ItemNumber=7194
insulin-like growth factor 1 (IGF-1) and the occurrence of FM syndrome. (133) Sodium oxybate increases growth hormone production levels through increased slow-wave sleep patterns. However, this medication was not approved by the FDA for the indication for use in people with FM due to the concern for abuse. (134, 135) The use of NSAIDs are not recommended as the first line therapy. (136) NSAIDs are not regarded to be useful in the management of FM. (137)

1.4.2. Non-pharmacological treatment

The non-pharmacological treatment together with the related therapies is an application of psychosomatic medicine to resolve the problem of FM. Cognitive behavioural therapies (CBTs) include interventions that are based on the basic premise that chronic pain is maintained by cognitive and behavioural factors, and that psychological treatment leads to changes in these factors through cognitive (e.g., cognitive restructuring) and/or behavioural (e.g., relaxation training, social skills training) techniques. Different types of CBTs can be differentiated by the techniques applied.

As a non-pharmacological component of the management of FM, CBTs and related therapies have a small to moderate effect in reducing FM symptoms. (138) CBTs have been recommended for FM management by recent evidence-based guidelines. (125, 139) Internet-based cognitive psychological therapies (ICBTs), a subset of CBTs, are a growing area of mental health, because they are cost-effective and almost all-pervading in many somatic and psychological disorders. (139)

A systematic review and meta-analysis of the efficacy of CBT in FM was performed, including 14 trials with 910 subjects with a median treatment time of 27 hours. Though the positive effect on depressed mood was not identified with possible risks of bias, there was a significant effect on self-efficacy pain post treatment, and the ‘healthcare-seeking behaviour’ was also reduced (subjects
reduced the frequency to see physicians at follow-up). (140)

In addition to CBTs, mind-body therapy is an increasing social and clinical trend, which focuses on interactions among the brain, mind, body and behaviour. (141) It is a synthesis of biofeedback, psychological and physical interventions. Psychological interventions are slightly effective in the treatment of FM, but the sustained effectiveness of mind-body therapy remains to be substantiated (142, 143)

There is strong evidence indicating that exercise improves fitness and sleep and may reduce pain and fatigue in some people with FM. (144, 145) Low-quality evidence from heterogenous trials suggests that high-intensity resistance training may improve pain and strength in women. (146) Studies of different forms of aerobic exercise for adults with FM imply that aerobic exercise improves quality of life, reduces pain, slightly improves physical function and yet makes no difference in fatigue and stiffness. (147) In terms of combinations of exercises, Bidonde et al investigated randomised controlled trials (RCTs) in adults with a diagnosis of FM that compared mixed exercise interventions with other or no exercise interventions. (148) They found that combinations of different exercises such as flexibility and aerobic training may improve stiffness, but the improvement may be of less clinical importance for some participants, and it remains uncertain whether the long-term effects can be maintained. Table 1-3 summarises different types of management of FM.

Table 1-3 Summary of management of FM

<table>
<thead>
<tr>
<th>Types of management</th>
<th>Examples</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication</td>
<td>Antidepressants</td>
<td>SSRIs (Duloxetine, milnacipran, etc.)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TCAs (amitriptyline, etc)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MAOIs (Pirlindole, moclobemide, etc.)</td>
</tr>
<tr>
<td>Anticonvulsants</td>
<td></td>
<td>Pregabalin, gabapentin, etc</td>
</tr>
</tbody>
</table>
### Table: Treatment Options

<table>
<thead>
<tr>
<th>Category</th>
<th>Therapy</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opioids</td>
<td></td>
<td>Effective but not recommended (with possibility of substance abuse)</td>
</tr>
<tr>
<td>Others</td>
<td>Sodium oxybate</td>
<td>Not recommended by FDA</td>
</tr>
<tr>
<td></td>
<td>NASIDs</td>
<td>Not useful</td>
</tr>
<tr>
<td>Therapy</td>
<td>CBTs</td>
<td>Cost-effective and evidence-based</td>
</tr>
<tr>
<td></td>
<td>ICBTs, Mindfulness-based stress reduction (MBSR), etc.</td>
<td></td>
</tr>
<tr>
<td>Exercise</td>
<td>Mind-body therapy</td>
<td>Slightly effective</td>
</tr>
<tr>
<td></td>
<td>Aerobic exercise, flexibility exercise, resistance exercise, etc.</td>
<td>Less clinically important to some patients; long-term effects not found</td>
</tr>
</tbody>
</table>

### 1.5. Introduction to the UK Biobank

The UK Biobank is a major national and international health resource as well as a registered charity. It is established with the aim of improving the prevention, diagnosis, and treatment of a wide range of diseases. Five hundred thousand 40-69-year-old people were recruited across the UK by the project from 2006 to 2010. All participants gave informed consent for their data to be used for research purposes. Samples of blood, urine and saliva are stored for future analysis. Details of the UK Biobank resource can be found at [www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk). Figure 1-5 shows the UK Biobank resource and genotyping array content. (149)
1.5.1. The UK Biobank Axiom genotyping array

The UK Biobank Axiom array from Affymetrix is a widely adopted platform for large-scale biobank genotyping studies. It was specifically designed for genotyping the UK Biobank participants. There are around 800,000 markers on the array. The design concept was basically categorized into three parts: (a) to add markers of particular interest based on known associations and/or possible roles in phenotypic variation; (b) to add coding variants across a range of minor allele frequencies (MAFs), principally missense and protein truncating variants; (3) to choose the remaining content to provide good genome-wide imputation coverage in European populations in the common (> 5%) and low frequency (1-5%) MAF ranges.
For GWA studies, Axiom array content modules are able to provide intelligent marker selection which enables imputation of millions of additional SNPs. The genomic content module of each array is also optimised for specific populations and is customisable to any 1000 Genomes population study by adding additional markers dependent on the specific research focus.

1.6. Objective of the study

The purpose of this study is to identify the genetic variants related to CWP ('pain all over the body'). We conducted a genome-wide association study (GWAS) using the cohort dataset from UK Biobank.
2. Methodology

2.1. Basics of GWAS

Genome-wide association study (GWA study, or GWAS), also known as whole genome association study (WGA study, or WGAS), aims to explore/examine associations between genetic variants (single nucleotide polymorphisms, SNPs) and major human diseases on a scale of whole genome set. This method is a type of observational study and entails the correspondence of a given human genome sequence with a fully interpreted map of common genetic variation. When a study focuses on a large population with well-defined clinical characteristics and a large collection of DNA samples is provided, GWAS is a powerful research tool. By sifting out relevant genes, GWAS is providing us with a new realm of understanding that individuals are predisposed towards certain diseases due to genetic variations.

GWAS is theoretically based on the linkage disequilibrium (LD), a non-random relation of alleles at different loci in a given population. In other words, SNP alleles or DNA sequences that are close together in the genome tend to be inherited together. Once the patterns of LD are known, alleles can play roles as markers, either individually or collectively, to distinguish one region from another, and thus fewer SNPs can capture more common variations within the regions. Besides, the possible SNP variants are so many that it is still very expensive to sequence all SNPs, and the feasible way is to set up customisable arrays to genotype a sub set of the variants. For instance, there are more than eight hundred thousand SNPs in the bespoke Affymetrix UK Biobank gene chips used in this study, and all variants of the genome can virtually be tagged. If a SNP is related to a trait or a disease, then the actual biological effect will be brought forth by one of the variants that the SNP stands for, including that SNP itself.

Since GWA studies aim to investigate the whole genome, it is a method not driven by any candidate genes, in contrast with other genetic methods that specifically
test a small number of pre-specified genetic regions. Candidate genes are most often selected for study based on *a priori* knowledge of the biological functional impact of gene on the trait or disease in question. (150) Candidate genes are often confined within a small number of genetic regions. If the previous research has flaws or fails to be replicated, then the candidate genes deduced from it will be questioned. GWAS can avoid that limitation. On the other hand, GWA studies cannot deduce on their own which genes are causal, while they are able to identify SNPs and other variants in DNA related to a certain disease.

Figure 2-1 General classification of GWA studies
2.1.1. GWAS based on unrelated individuals

There are two types of designs when targeting unrelated individuals (see Figure 2-1): case-control studies and population-based association studies. The former is mainly used to study qualitative traits (whether it is diseased) while the latter is generally to discover quantitative traits. Various statistical approaches are used, which depends on the study design and the phenotype. For instance, a case-control study design (quality traits), comparing the allele frequency of each SNP between the case and the control group can be performed using a 2 x 2 chi-squared test to calculate the odds ratio (OR) and its 95% confidence interval (CI). The attributable fraction (AF) as well as attributive risk (AR) can also be calculated. Logistic regression is also used when adjustments are made to avoid confounding factors e.g., age, sex, and ethnicity. The general process is to regard the occurrences of diseases as the dependent variable, and the genotypes as well as confounding factors as independent variables. The study design is often based on a random population (quantitative traits), such as investigating the association between SNPs and the quantitative phenotypes of a disease (BMI, etc.). It is common to compare whether the levels of the phenotype among the genotype carriers at the locus are different (one-way ANOVA). Confounding factors probably need to be adjusted by introducing covariance analysis or linear regression equations.

2.1.2. Multiple test adjustment in GWA studies

The statistical issue of multiple testing is one of the greatest problems that contribute to potential false-positive results because of the massive numbers of SNPs. (151) The common methods for multiple test corrections are generally consolidated into the software for genetic statistics (see Figure 2-2).
The Bonferroni correction compensates for the increasing likelihood of rejecting a null hypothesis incorrectly (i.e., making a Type I error) by testing each individual hypothesis at a significance level of $\frac{\alpha}{m}$, where $\alpha$ represents the desired overall alpha level and $m$ represents the number of hypotheses. (152) The Bonferroni correction receives criticism of the conservative trait since there are too many times of comparisons and the test statistics are positively correlated. The false-negative rate is usually high.

The Holm-Bonferroni method is intended to control the family-wise error rate (FWER) and offers a simple test uniformly more powerful than the Bonferroni correction. (153, 154) The basic procedure is as follows: By ordering corresponding $p$-values from lowest to highest ($p_1, p_2, ..., p_m$), the associated hypotheses are assumed to be $H_1, H_2, ..., H_m$. For a given significance level $\alpha$, let $k$ be the minimal index such that $p_k > \frac{\alpha}{m-k+1}$. Then reject the null hypotheses from $H_1$ to $H_{k-1}$ and do not reject the rest ($H_k, H_{k+1}, ..., H_m$). Specially, all hypotheses are rejected if $k = 1$. The method ensures that FWER is not more than $\alpha$. Like Bonferroni correction, the Holm-Bonferroni method adjusts the $p$-values separately. It requires software and advanced methods to calibrate $p$-values simultaneously.

The Westfall-Young permutation takes strongly dependent test-statistics into account. (155) The basic operation is to sort out the uncorrected $p$-values at first, and then repeat sampling simulations based on the structural relationship
between genes. After analysing the distribution of \( p \)-values, correct all \( p \)-values simultaneously.

The Benjamini-Hochberg procedure aims to decrease the false discovery rate and control the \( p \)-values still further. (156) The basic procedure is as follows: By ordering corresponding \( p \)-values from lowest to highest, keep the largest \( p \)-value and multiply others by certain coefficients (amount of loci/order of the \( p \)-value). If the adjusted \( p \)-value is smaller than the given significance level \( \alpha \), then the locus makes a significant contribution to the disease. The Benjamini-Hochberg procedure is one of the least strict corrections of \( p \)-values, which would cause more false positive cases.

2.1.3. Limitations in GWAS

GWAS opens a new chapter in the study of complex diseases: researchers do not need to presuppose any strategies for candidate genes, but instead compare the allele frequency of all the genome-wide mutations in cases and controls and find disease-related sequence variation. GWAS has discovered many unknown genes and chromosomal regions that we have not known previously, providing more clues for us to understand the pathogenesis of complex human diseases. However, we cannot be too optimistic about the role of GWAS in the aetiological studies of complex diseases. To find SNPs which are truly related to complex diseases, GWAS requires the following conditions: the case group of the GWAS sample must carry the genetic factors resulting in the disease; the research needs to achieve a sufficient test power; the sample size and the number of SNPs are very large; a large number of human SNPs need to be genotyped in thousands of cases and controls. Data analysis also requires more advanced statistical methods. (151)

Sometimes, even if the GWAS results suggest a disease-associated chromosomal region, it is still difficult to determine the true disease-causing SNP. The reason is that the functional SNPs that cause disease occur within a gene with a
large degree of variability, either in the coding DNA sequence, the splicing site, or in the regulatory region of a gene. Even though the association with disease found by GWAS is validated, only a few currently believe that these results can be used to guide clinical practice in the short term, such as to assess a person’s risk of developing a disease. (157) Although GWAS has found many SNPs that may be related to disease phenotypes, it is still unclear how these genes interact with environmental factors and how lifestyle changes regulate the roles of these genes.

2.2. Participants and their genetic information

The UK Biobank cohort recruited over 500,000 people aged between 40 and 69 years in 2006–2010 across the UK. Participants completed a detailed clinical, demographic, and lifestyle questionnaire, underwent clinical measures, provided biological samples (blood, urine and saliva) for future analysis, and agreed to have their health records accessed. The informed consent of all participants has been obtained. Details of the UK Biobank resource can be found at www.ukbiobank.ac.uk. UK Biobank received ethical approval from the National Health Service National Research Ethics Service (reference 11/NW/0382). The current analyses were conducted under approved UK Biobank data application number 4844. The details of genotype quality, properties of population structure and relatedness of the genetic data, and efficient phasing and genotype imputation are described by Bycroft, C. et al. (149)

2.3. Phenotypic information on pain

The 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. These body-site pain options were not mutually exclusive, and
participants could choose as many as they felt appropriate. The pain all over the body cases in this study were those who selected the ‘pain all over the body’ option in response to the question, regardless of whether they had selected other options. The controls in this study were those who selected the ‘None of the above’ option. Thus, the same ‘no pain’ control population was used for all pain phenotypes in different body sites.

2.4. Statistical analysis

The source material was derived from the raw data from the UK Biobank and underwent GWAS approaches by BGENIE. (158) In this study, we used BGENIE (https://jmarchini.org/bgenie/) to be the main GWAS software and removed single nucleotide polymorphism (SNPs) with INFO scores < 0.1, with minor allele frequency < 0.5%, or those that failed Hardy-Weinberg tests \( P < 10^{-6} \). The GWAS model was a mixed linear regression model, and the genetic model was additive as default. SNPs on sex chromosomes and the mitochondrion as well as imputed SNPs that were not in the Haplotype Reference Consortium panel were excluded from analyses. Standard Frequentist association tests using BGENIE was used to perform association studies adjusting for age, sex, body mass index (BMI), 9 population principal components, genotyping arrays, and assessment centres. The gender difference between cases and controls was compared using chi-square testing. Age and BMI were compared using independent t testing in IBM SPSS 24 (IBM Corporation, New York). SNP associations were considered significant if they had a \( p \) value < \( 5 \times 10^{-8} \). GCTA was used to calculate SNP-based or narrow-sense heritability using a genomic relationship matrix calculated from genotyped autosomal SNPs.

The main part of the study was to run the GWAS summary on FUMA (Functional Mapping and Annotation of Genome-Wide Association Studies, see Figure 2-4), a platform that can be used to annotate, prioritize, visualize and interpret GWAS results. (159) Compared to other applications for SNP annotation e.g., GWAS4D,
SNPnexus, etc., the advantages of FUMA are the consolidation of multifunctional mainstream resources of GWAS, continuous updates and timely maintenances, and less burden of offline calculation. The operation of FUMA is also simplified, for most of the calculations are automatically executed after uploading and submitting the GWAS file.

Figure 2-4 Flowsheet of the methodology

The SNP2GENE function takes GWAS summary statistics as an input and provides extensive functional annotation for all SNPs in genomic areas identified by lead SNPs. To characterise the significant hits, the SNP2GENE can be broken down into three steps (see Figure 2-5). The first step is to characterise genomic loci, including identifying independent significant SNPs and candidate SNPs, defining leading SNPs, and genomic risk loci. The second step is to annotate candidate SNPs in genomic loci, with tests of expression quantitative trait loci (eQTL), 15 chromatine state, 3D chromatine interactions (Hi-C), etc. The third step is to generate the table of functional gene mapping, which is used in GENE2FUNC, another main process of FUMA. The interact visualisation (Manhattan plot, etc) is based on the results from all three steps, with p-values calculated by MAGMA.

---

5 From https://fuma.ctglab.nl/tutorial
(Multi-marker Analysis of GenoMic Annotation, integrated in FUMA) gene analysis and MAGMA gene-set analysis (GSA). GSA makes genes be aggregated into gene sets on the basis of shared biological or functional properties according to certain knowledge bases. Knowledge bases are database collections of molecular knowledge that may encompass molecular interactions, regulation, molecular products and even phenotype associations. The resultant gene sets are analysed as a whole to determine which of these properties are relevant to the phenotype of interest. GSA is able to generate hypotheses on phenomenalistic processes for the phenotypes of interest, and the replication and laboratory experiments should also be done to validate the results. (160, 161) The MHC region was excluded from the analysis, and SNPs in that region were not annotated.

Figure 2-5 Basic functions of SNP2GENE of FUMA
3. Results

3.1. GWAS results

There were 872,316 items of completed records in response to the question of ‘pain type(s) experienced last month’ in the questionnaire, covering 501,600 patients. Among these patients, 9,551 reported having experienced activity-limiting pain all over the body in the previous month. There were 232,467 participants who selected ‘none of the above’ option, which meant that they did not have any form of pain that limited the activity in the previous month.

Firstly, samples were checked to establish a homogeneous dataset. Ancestry information provided from the UK Biobank was used in the process. In addition, those who were related to one or more in the cohort (a cut-off value of 0.044 (162, 163) in the generation of the genetic relationship matrix) and those who failed quality control were also removed. After the exclusions (see Figure 3-1), the number in the case group fell to 5,670 (2,171 males vs 3,499 females, see Figure 3-1). Meanwhile, the control group encompassed 149,312 (71,480 males, 77,832 females) individuals. After the quality control of SNPs, 9,304,965 SNPs were available for the GWAS. Clinical characteristics of the cases and controls were compiled (Table 3-1). Age, sex, and BMI were all found to be significantly different ($p < 0.01$) between cases and controls.

A quantile-quantile plot (Q-Q plot) can be used to characterise the extent to which the observed distribution of the test statistic follows the expected distribution. In GWAS, the null hypothesis is that none of the SNPs in the test is related to the certain disease. With the null hypothesis, the P values from the tests where no true association exists should follow a continuous uniform distribution. The deviation at the tail of the Q-Q plot can deny the null hypothesis. Q-Q plot can be a useful a tool to evaluate differences between cases and controls caused by potential confounders.
With a threshold of $p < 5 \times 10^{-8}$, two genetic loci including three significant and independent SNPs were found in the GWAS (Figure 3-2, Table 3-2). The most significant SNP (rs550883786, $p = 1.69 \times 10^{-8}$) was found in an intergenic region in chromosome 4. The first locus was found in the FAF1 gene located in chromosome 1, and the most significant SNP from this locus was rs17387024 ($p = 3.72 \times 10^{-8}$). The second locus was the LINC01078 (long intergenic non-protein coding RNA 1078) gene located in chromosome 13, and the most significant SNP in this locus was rs148500993 ($p = 3.65 \times 10^{-8}$). The highest peak shown in the Manhattan plot was in chromosome 6, but there were no significant SNPs annotated in that area, and no SNPs in chromosome 6 were recorded in FUMA results files. Due to the fact that the LD among SNPs was very complicated in this area, FUMA does not annotate this region by default as most of the FUMA related analyses (such as MAGMA) was based on the correctness of LD. (159) Besides, no publications showed any direct association between the MHC region and FM/CWP.
Figure 3-1 Original & adjusted statistics of cases & controls in the UK Biobank

![Bar chart showing original and adjusted statistics for cases and controls in the UK Biobank.]

Table 3-1 Genders, ages and BMIs of cases and controls of CWP

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Cases</th>
<th>Controls</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (male:female)</td>
<td>2,171(38.3%):</td>
<td>71,480(47.9%):</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>3,499(61.7%)</td>
<td>77,832(52.1%)</td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td>57.3 (7.59)</td>
<td>56.9 (7.97)</td>
<td>0.002</td>
</tr>
<tr>
<td>BMI (Body mass index, kg/m²)</td>
<td>29.6 (5.96)</td>
<td>26.7 (4.30)</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

The chi-square test was used for gender frequency between cases and controls, while Student t-tests were used for others. Continuous covariates were presented as mean (standard deviation).
Figure 3-2 The Manhattan plot of the GWAS on CWP using the UK Biobank
<table>
<thead>
<tr>
<th>rsID</th>
<th>chr</th>
<th>position</th>
<th>start</th>
<th>end</th>
<th>p-value (10^-8)</th>
<th>allele1:allele2</th>
<th>nSNPs</th>
<th>nGWASSNPs</th>
<th>Related gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs17387024</td>
<td>1</td>
<td>50993178</td>
<td>50993178</td>
<td>51476972</td>
<td>3.72</td>
<td>C:T</td>
<td>6</td>
<td>6</td>
<td>FAF1</td>
</tr>
<tr>
<td>rs550883786</td>
<td>4</td>
<td>29485633</td>
<td>28609064</td>
<td>30094647</td>
<td>1.69</td>
<td>C:T</td>
<td>12</td>
<td>1</td>
<td>(intergenic region)</td>
</tr>
<tr>
<td>rs148500993</td>
<td>13</td>
<td>75823871</td>
<td>74987745</td>
<td>76754250</td>
<td>3.65</td>
<td>C:G</td>
<td>9</td>
<td>2</td>
<td>LINC01078</td>
</tr>
</tbody>
</table>

**Table 3-3** The significant gene sets with Bonferroni p value < 0.05.

<table>
<thead>
<tr>
<th>Standard gene-set name</th>
<th>Systematic name</th>
<th>Category</th>
<th>Number of genes</th>
<th>β</th>
<th>β SD</th>
<th>p-value (×10^-6)</th>
<th>Bonferroni p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>davicioni_targets_of_pax_foxo1_fusions_up</td>
<td>M4680</td>
<td>Curated</td>
<td>243</td>
<td>0.283</td>
<td>0.032</td>
<td>1.29</td>
<td>0.0200</td>
</tr>
<tr>
<td>missiaglia_regulated_by_methylation_up</td>
<td>M12272</td>
<td>Curated</td>
<td>112</td>
<td>0.356</td>
<td>0.027</td>
<td>2.33</td>
<td>0.0360</td>
</tr>
<tr>
<td>go_structural_constituent_of_synapse</td>
<td>M26829</td>
<td>Ontology</td>
<td>14</td>
<td>1.072</td>
<td>0.029</td>
<td>2.67</td>
<td>0.0413</td>
</tr>
<tr>
<td>go_presynaptic_cytoskeleton</td>
<td>M26024</td>
<td>Ontology</td>
<td>8</td>
<td>1.364</td>
<td>0.028</td>
<td>3.21</td>
<td>0.0496</td>
</tr>
</tbody>
</table>
Figure 3-3 The Q-Q plot of GWAS summary statistics.

Figure 3-4 The Q-Q plot of the gene-based test computed by MAGMA.

For plotting purposes, overlapping data points are not drawn. Lambda = 1.09.
The Q-Q plots (Figure 3-3 and 3-4) were generated to complement the results shown in Manhattan plot. The observed $p$-values at the extreme deviated from the expected null distribution (uniform distribution), which suggested possible un-adjusted covariates between cases and controls.

### 3.2. FUMA results

In gene-based association analysis by MAGMA (integrated into FUMA), a total of 10,894 gene sets were tested and a default competitive test model was applied. Eight genes were found to be associated with CWP (Figures 3-4 & 3-5). The most significant gene was MRPL44 ($p = 1.30 \times 10^{-8}$) located at chromosome 2. The genes found in the gene-based test were located in completely different chromosomes compared with the significant and independent SNPs found in GWAS.

In gene-set analysis conducted by MAGMA, 15,496 gene-sets were analysed using the default competitive test model. Gene sets of positive regulation of gene expression, positive regulation of transcription from RNA polymerase II promoter, neurogenesis, and excitatory synapse reached a $P$ value $< 0.0001$, but not statistically significant of $p < 5 \times 10^{-6}$ ($0.05/10,894$). The top four gene-sets met genome-wide significance and were included in Table 3-3.

Tissue expression analysis was conducted by GTEx (Genotype-Tissue Expression), and the relationship between tissue-specific gene expression and genetic associations was tested by using the average gene expression in each certain tissue type as a covariate. Two analyses were carried out. One investigated 30 general tissue types (Figure 3-6) and the other focused on 53 specific tissue types (Figure 3-7). Tissue expression analysis in 30 tissue types did not find significant expression in any of the tissue types. Tissue expression analysis of 53 specific tissue types by GTEx found that the expression in the cerebellar hemisphere had the lowest $p$ value ($p = 1.03 \times 10^{-3}$), and the expressions in other parts of brain
such as cerebellum, cortex, and frontal cortex BA9 also had relatively low p-values, but none of them reached the significant $p$ value, either.
Figure 3-5 The Manhattan plot of the gene-based test as computed by MAGMA based on the input GWAS summary statistics\(^6\)

\(^6\) Genome wide significance (red dashed line in the plot) is defined at \(p = 0.05/19043 = 2.626 \times 10^{-6}\).
Figure 3-6 Tissue expression results on 30 specific tissue types by GTEx in the FUMA.

The blue columns indicate that they do not reach the cut-off $p$-value for significance with Bonferroni adjustment for multiple hypothesis testing.
Figure 3-7 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.

10 The dashed line shows the cut-off p-value for significance with Bonferroni adjustment for multiple hypothesis testing.
4. Discussion

4.1. Lead SNPs and related genomic loci

We have derived three loci which reached the genome-wide significance (p < 5×10^{-8}) from this GWAS on pain over the body with the resources from the UK Biobank. They are respectively located at the FAF1 gene in chromosome 1, an intergenic region in chromosome 4 and the LINC01078 gene in chromosome 13.

In this study, we defined the term ‘pain all over the body’ based on the responses of the participants from the UK Biobank. The generic pain question generated by the UK Biobank is a useful implement for testing whether a heterogeneous pain phenotype (e.g., neck and shoulder pain, knee pain, pain all over the body, etc.) has genetic components in nature or not. For instance, Meng et al. used the same question as an identification of the genetic variants of broadly defined headache, and their findings basically tallied with other researches in well-defined migraine phenotypes. (164, 165) One of the advantages to use the UK Biobank on heterogeneous phenotypes is that researchers are able to use large numbers to reduce statistical noise, and they can compensate for potentially reduced power because of heterogeneity.

The top locus is at the intergenic region in chromosome 4 with the lowest p value of 1.69×10^{-8} for rs550883786, and there is one lead SNP as well as 11 other unique candidate SNPs (known as nIndSigSNPs) in the genomic locus. The second locus is at the LINC01078 gene in chromosome 13 with the second lowest p value of 3.65×10^{-8} for rs17387024, and there are two GWAS-tagged candidate SNPs out of nine unique candidate SNPs in the genomic locus. The LINC01078 gene is a gene with 441 nt (nucleotides) which codes RNA 1078 rather than any sorts of proteins^{11}. There have been no publications about this gene to date. Its nearest gene which encodes a protein is the TBC1D4 gene (see Figure 4-1), with

---

^{11} See https://www.genecards.org/cgi-bin/carddisp.pl?gene=LINC01078
a length of 199,642 nt. TBC1D4, also named as AS160 (Akt substrate of 160 kDa), is the Rab-GTPase-activating protein. Researches have revealed that TBC1D4 plays an important role in the homoeostasis of glucose with the regulation of insulin-dependent trafficking of the glucose transport type 4 (GLUT-4, also known as solute carrier family 2), which helps to remove glucose from the bloodstream to fat tissues and skeletal muscle. (166, 167) When exposed to insulin, this protein is phosphorylated, dissociates from GLUT4 vesicles, resulting in increased GLUT4 at the cell surface, and enhanced glucose transport. In spite of the pivotal role of TBC1D4 in the pathophysiology of diabetes, the features, and functions of TBC1D4 seem to be irrelevant to the mechanisms of either FM or CWP, and no connection can be established between the TBC1D4 gene and the phenotype ‘pain all over the body’.

Figure 4-1 Locations of LINC01078 and TBC1D4 gene in chromosome 13

Figure 4-2 Location of FAF1 gene in chromosome 1

The final locus is at the FAF1 gene in chromosome 1p32.3 (see figure 4-2) with the third lowest p value of 3.72×10⁻⁸ for rs17387024. rs17387024 is the lead SNPs in the genomic locus, while the number of GWAS-tagged SNPs which are in LD of the independent significant SNPs given r² (known as nGWASSNPs) is six. The length of FAF1 gene is 523,240 nt and the gene encodes the protein

---

12 Screenshot from https://www.ncbi.nlm.nih.gov/snp/rs148500993
13 Screenshot from https://www.genecards.org/cgi-bin/carddisp.pl?gene=FAF1
named fas-association factor 1 (FAF1). The protein contains 650 amino acids with a mass of 73,954 Da. Initially identified as a Fas-binding protein, FAF1 was found to increase the likelihood of Fas-induced apoptosis. The functions of FAF1 are various and it takes part in multiple mechanisms which promote cell death (see Table 4-2). It exerts influence on both extrinsic and intrinsic pathways and mediates the activation of caspase 8. FAF1 can also suppress the activation of NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells) by interfering IKK (IkB kinase) complex assembly, and the lowered complex formation is related with physical interaction with IKKβ. The negative feedback regulation of Aurora-A via phosphorylation of FAF1 is also supported, which clarifies one of the intermediate links of cell cycle. Besides, the interactions between FAF1 and VCP (valosin-containing protein, as known as transitional endoplasmic reticulum adenosine triphosphatase, TER ATPase) have been found to inhibit the ubiquitin dependent protein degradation. The downregulation of FAF1 has been observed in gastric and uterine cervix carcinomata, and thus it is thought to function as one of the tumour suppressors through the activities in promotion of apoptosis.

Along with its suppressive roles in the development of tumour, FAF1 is found to be involved in the pathogenesis of Parkinson’s disease (PD). Betarbet et al. found a significant increase of FAF1 expression in the prefrontal cortices of patients with PD, and FAF1 enlarges the toxic effects of stressors related to PD, such as oxidative stress. Furthermore, as a pathogenic substrate of parkin, a ubiquitin E3 ligase, FAF1 plays a determining role in the dopaminergic neural degeneration: the inactivation of parkin by PD-linked mutations or by genetic deletion causes the uplifted expression FAF1 and other relevant biochemical reactions mediated by it. The influenced events in animal models encompass caspase activation, ROS (reactive oxygen species) generation, JNK (c-Jun N-terminal kinase) activation and cell death. The related experiments imply that the pathogenesis of PD and the oxidative stress-induced cell death pivot on the action of FAF1 on
the apoptotic mechanism. FAF1 is also involved in the pathogenesis of ischemic
diseases. Yu et al. established a retinal ischemia model and found that FAF1 was
a key factor of JNK1-dependent necrosis upon ischemic stress, which was fol-
lowed by mitochondrial dysregulation. (177)

The finding of the FAF1 gene that is genetically associated with the phenotype
‘pain all over the body’ coincides with another GWAS of multisite chronic pain
(MCP) in the UK Biobank. (94) Apart from FAF1 gene, the study also found that
38 other genomic risk loci, and 76 independent genome-wide significant SNPs
associated with MCP were identified in total, which was a lot more fruitful. Alr-
hough MCP, CWP, and pain all over the body are all subsets of chronic pain,
these phenotypes are slightly different by definition. MCP stresses the number of
sites at which chronic pain is experienced. According to the UK Biobank ques-
tionnaire (field ID 6159), seven individual body-site pain options were not mutu-
ally exclusive, and participants could choose more than one answer, except for
those who chose ‘pain all over the body’, ‘none of the above’ or ‘prefer not to say’.
MCP was defined as the sum of body sites at which chronic pain (at least 3
months duration) was recorded: 0 to 7 sites, based on an additional question in
UK Biobank (Category 100048). Those who answered that they had chronic pain
‘all over the body’ were excluded from the GWAS.
<table>
<thead>
<tr>
<th>Interaction(s)</th>
<th>Function(s)</th>
<th>Method(s) of detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNFRSF6 (tumour necrosis factor receptor superfamily member 6)</td>
<td>Regulation of apoptosis</td>
<td>Yeast two-hybrid screening and overexpression in mouse L cells; overexpression in human BOSC23 cells (168, 169)</td>
</tr>
<tr>
<td>Fas-associated protein with death domain (FADD/MORT1) and caspase 8</td>
<td></td>
<td>Co-immunoprecipitation, co-localization, and overexpression in Jurkat cells (169)</td>
</tr>
<tr>
<td>Protein kinase casein kinase 2 beta (CK2beta)</td>
<td></td>
<td>Yeast two-hybrid; immunoprecipitation and co-localisation studies (178, 179)</td>
</tr>
<tr>
<td>nuclear factor NF-kappa-B p65 subunit</td>
<td>Inhibition of NF-κB by cytoplasmic retention of p65</td>
<td>Overexpression and immunoprecipitation studies in HEK-293 and NIH3T3 cells (180)</td>
</tr>
<tr>
<td>IKK (IκB kinase)</td>
<td>Suppression of IKK activation linked to regulation of NF-κB</td>
<td>Immunoprecipitation, overexpression and siRNA studies in HEK293 cells (170)</td>
</tr>
<tr>
<td>Pyrin domains of pyrin domain-containing Apaf1-like proteins (PYPAF5s)</td>
<td>Inflammatory signalling associated with NF-κB</td>
<td>Yeast two-hybrid screen, immunoprecipitation, and co-expression in HEK 293 and HEK293T cells (181)</td>
</tr>
<tr>
<td>Ubiquitinated proteins and valosin-containing protein (VCP)</td>
<td>Regulation of protein degradation in the ubiquitin-proteasome pathway</td>
<td>Overexpression in HEK293T cells; protein dissociation of rat skeletal muscle and mass spectrometry (108, 172)</td>
</tr>
<tr>
<td>Heat shock protein 70 (Hsp70)</td>
<td>Inhibition of Hsp70 chaperone of refolding denatured protein substrates</td>
<td>Co-immunoprecipitation, peptide mass fingerprinting, and co-localization in HEK293T cells (182)</td>
</tr>
<tr>
<td>Mineralocorticoid receptor (MR) and</td>
<td>Modulates transactivation potential; selectively</td>
<td>Yeast two-hybrid screening and co-expression in HN9.10 mouse</td>
</tr>
<tr>
<td>Interaction(s)</td>
<td>Function(s)</td>
<td>Method(s) of detection</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>--------------------------------------------------</td>
<td>------------------------------------------------------------</td>
</tr>
<tr>
<td>glucocorticoid receptor (GR)</td>
<td>stimulates MR-mediated transcription</td>
<td>hippocampal cells (183)</td>
</tr>
<tr>
<td>Transient receptor potential vanilloid 1 receptor (TRPV1)</td>
<td>Modulates sensitivity to noxious stimuli (capsaicin, acid, heat) in sensory neurons</td>
<td>Co-expression and immunoprecipitation from sensory neurons or HEK293T cells (184)</td>
</tr>
</tbody>
</table>
4.2. Gene-set analysis

Four gene-sets were found significant according to the analysis by MAGMA. However, there is little connection between the found gene-sets and FM/CWP. The most significant gene set M4680 was briefly described as Genes up-regulated in RD cells (embryonal rhabdomyosarcoma, ERMS) by expression of PAX3- or PAX7-FOXO1 [GenetID=5077;5081;2308] fusions off retroviral vectors. (185) Gene set M12272 was briefly described as genes up-regulated in PaCa44 and CFPAC1 cells (pancreatic cancer) after treatment with decitabine [PubChem=451668], a DNA hypomethylating agent similar to azacitidine [PubChem=9444]. (186) The other two gene-sets are lack of enough publications to reveal their relationship with FM/CWP.

4.3. Limitations in the study

It could be a drawback that the phenotype was based on answers by participants’ self-reporting. However, the sample size was large enough so it should have sufficient power to identify genetic factors related to pain all over the body. Although the observational studies have shown that females are more likely to have symptoms of FM compared to males, the study did not include any genes on sex chromosomes. Thus, the genomic relationship between gender and FM/CWP was not able to be revealed. It could also cause biases because of different criteria of timespan, as the questionnaire is concerned about only one month.

According to the Manhattan plot, the most significant SNP should have been located in chromosome 6, but the algorithm excluded the MHC region and no significant SNPs in chromosome 6 were included in the output, which could be another limitation.

The Q-Q plot suggested that there could be residual confounding factors between cases of pain all over the body and controls that have not been adjusted for.
Replication of the study is also difficult owing to the fact that the nature of chronic pain phenotyping is not unanimous and available cohort sizes are not always consistent. Nevertheless, there are resemblances among identified genomic risk loci in other GWA studies in complex traits pertaining to chronic pain, including chronic back pain, migraine, and FM, etc. There was only one gene of interest found in the GWAS and no significant results in the tissue expression analysis were shown.

5. Conclusion

Pain all over the body is one of the subsets of pain, and it is the complement to other phenotypes of pain in specific regions (headache, knee pain, neck and shoulder pain, etc). The study revealed part of the genomic traits of this phenotype by using the UK Biobank. It is the first research to explore the genetic factors of pain all over the body. The study suggested that the FAF1 gene could be related to pain all over the body, along with cell-cycle and programmed cell death, and the expression was primarily within brain tissues. The results conformed with other genetic and neurological chronic pain studies in which functional and structural alternations of the brain were found to contribute to chronic pain. (187, 188) This study could be a valuable supplement to the studies in other pain phenotypes in the UK Biobank. The study may also compare with further studies in pain phenotypes with similar definitions (e.g., multisite chronic pain, widespread chronic pain, etc.). Further studies should also encompass sex chromosomes to uncover the genomic relationship between sex and fibromyalgia/chronic widespread pain. The loci we suggested could act as potential drug targets for further investigation.
References


45. Harris RE, Clauw DJ, Scott DJ, McLean SA, Gracely RH, Zubieta JK.
64


57. Liu YT, Shao YW, Yen CT, Shaw FZ. Acid-induced hyperalgesia and anxiodepressive comorbidity in rats. Physiol Behav. 2014;131:105-10.


85. Bouhouche A, Benomar A, Bouslam N, Chkili T, Yahyaou M. Mutation in the epsilon subunit of the cytosolic chaperonin-containing t-complex peptide-1 (Cct5)


107. Becker S, Schweinhardt P. Dysfunctional neurotransmitter systems in
fibromyalgia, their role in central stress circuitry and pharmacological actions on these systems. Pain Res Treat. 2012;2012:741746.


130. Codagnone MG, Podesta MF, Uccelli NA, Reines A. Differential Local


140. Bernardy K, Fuber N, Kollner V, Hauser W. Efficacy of cognitive-behavioral therapies in fibromyalgia syndrome - a systematic review and


### Appendix 1
All SNPs in LD with any of independent lead SNPs with r2 greater or equal to the defined threshold (part 1)

<table>
<thead>
<tr>
<th>uniqID</th>
<th>rsID</th>
<th>chr</th>
<th>pos</th>
<th>non-effect allele</th>
<th>effect allele</th>
<th>MAF</th>
<th>gwasP (10^{-8})</th>
<th>beta (10^{-3})</th>
<th>SE (10^{-4})</th>
<th>r2</th>
<th>IndSigSNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:50993178:C:T</td>
<td>rs17387024</td>
<td>1</td>
<td>50993178</td>
<td>C</td>
<td>T</td>
<td>0.1481</td>
<td>3.71706366796</td>
<td>5.1885</td>
<td>9.4266</td>
<td>1</td>
<td>rs17387024</td>
</tr>
<tr>
<td>1:51128777:A:G</td>
<td>rs10493152</td>
<td>1</td>
<td>51128777</td>
<td>G</td>
<td>A</td>
<td>0.17</td>
<td>15.8379876507</td>
<td>4.6495</td>
<td>8.8684</td>
<td>0.82505</td>
<td>rs17387024</td>
</tr>
<tr>
<td>1:51142556:C:T</td>
<td>rs17384367</td>
<td>1</td>
<td>51142556</td>
<td>C</td>
<td>T</td>
<td>0.09841</td>
<td>5.67022118161</td>
<td>6.1356</td>
<td>11.301</td>
<td>0.617948</td>
<td>rs17387024</td>
</tr>
<tr>
<td>1:51161263:A:G</td>
<td>rs78884406</td>
<td>1</td>
<td>51161263</td>
<td>A</td>
<td>G</td>
<td>0.1481</td>
<td>13.2342702837</td>
<td>4.8930</td>
<td>9.2744</td>
<td>0.967178</td>
<td>rs17387024</td>
</tr>
<tr>
<td>1:51264933:C:T</td>
<td>rs78091876</td>
<td>1</td>
<td>51264933</td>
<td>C</td>
<td>T</td>
<td>0.1511</td>
<td>25.4448561515</td>
<td>4.8493</td>
<td>9.4077</td>
<td>0.959427</td>
<td>rs17387024</td>
</tr>
<tr>
<td>1:51476972:C:T</td>
<td>rs77094699</td>
<td>1</td>
<td>51476972</td>
<td>T</td>
<td>C</td>
<td>0.1441</td>
<td>251.420103016</td>
<td>4.6284</td>
<td>9.8326</td>
<td>0.870554</td>
<td>rs17387024</td>
</tr>
<tr>
<td>4:28609064:A:C</td>
<td>rs573680863</td>
<td>4</td>
<td>28609064</td>
<td>C</td>
<td>A</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>4:28641266:A:T</td>
<td>rs529928586</td>
<td>4</td>
<td>28641266</td>
<td>T</td>
<td>A</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>4:28705788:C:G</td>
<td>rs51292161</td>
<td>4</td>
<td>28705788</td>
<td>C</td>
<td>G</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>4:28936170:A:C</td>
<td>rs530516790</td>
<td>4</td>
<td>28936170</td>
<td>A</td>
<td>C</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>4:28937341:C:T</td>
<td>rs567598400</td>
<td>4</td>
<td>28937341</td>
<td>C</td>
<td>T</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>4:29047725:A:C</td>
<td>rs11223043</td>
<td>4</td>
<td>29047725</td>
<td>A</td>
<td>C</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>4:29094291:C:T</td>
<td>rs535479714</td>
<td>4</td>
<td>29094291</td>
<td>C</td>
<td>T</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>4:29096073:C:G</td>
<td>rs571597976</td>
<td>4</td>
<td>29096073</td>
<td>C</td>
<td>G</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>4:29436739:C:T</td>
<td>rs531324916</td>
<td>4</td>
<td>29436739</td>
<td>C</td>
<td>T</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>4:29485633:C:T</td>
<td>rs550883786</td>
<td>4</td>
<td>29485633</td>
<td>T</td>
<td>C</td>
<td>0.000994</td>
<td>1.68655302539</td>
<td>82.379</td>
<td>0.014602</td>
<td>1</td>
<td>rs550883786</td>
</tr>
<tr>
<td>4:29807162:A:G</td>
<td>rs571291698</td>
<td>4</td>
<td>29807162</td>
<td>A</td>
<td>G</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>4:30094647:C:G</td>
<td>rs184775812</td>
<td>4</td>
<td>30094647</td>
<td>C</td>
<td>G</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>13:74987745:C:T</td>
<td>rs184775812</td>
<td>13</td>
<td>74987745</td>
<td>C</td>
<td>T</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>13:75032941:G:T</td>
<td>rs566173881</td>
<td>13</td>
<td>75032941</td>
<td>G</td>
<td>T</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>uniqID</td>
<td>rsID</td>
<td>chr</td>
<td>pos</td>
<td>non-effect allele</td>
<td>effect allele</td>
<td>MAF</td>
<td>gwasP (10^-8)</td>
<td>beta (10^-3)</td>
<td>SE (10^-4)</td>
<td>r2</td>
<td>IndSigSNP</td>
</tr>
<tr>
<td>------------</td>
<td>------------------</td>
<td>-----</td>
<td>-------</td>
<td>-------------------</td>
<td>---------------</td>
<td>---------</td>
<td>--------------</td>
<td>--------------</td>
<td>------------</td>
<td>---------</td>
<td>----------------------</td>
</tr>
<tr>
<td>13:75185289:C:T</td>
<td>rs564356221</td>
<td>13</td>
<td>75185289</td>
<td>T</td>
<td>C</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>rs148500993</td>
</tr>
<tr>
<td>13:75220378:C:T</td>
<td>rs148567855</td>
<td>13</td>
<td>75220378</td>
<td>C</td>
<td>T</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>rs550883786</td>
</tr>
<tr>
<td>13:75288942:A:G</td>
<td>rs542414084</td>
<td>13</td>
<td>75288942</td>
<td>G</td>
<td>A</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>rs148500993</td>
</tr>
<tr>
<td>13:75477615:G:T</td>
<td>rs563178847</td>
<td>13</td>
<td>75477615</td>
<td>T</td>
<td>G</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>rs148500993</td>
</tr>
<tr>
<td>13:75823871:C:G</td>
<td>rs148500993</td>
<td>13</td>
<td>75823871</td>
<td>C</td>
<td>G</td>
<td>0.000994</td>
<td>3.64753946926</td>
<td>63.324</td>
<td>114.98</td>
<td>1</td>
<td>rs148500993</td>
</tr>
<tr>
<td>13:75873067:G:T</td>
<td>rs562398442</td>
<td>13</td>
<td>75873067</td>
<td>T</td>
<td>G</td>
<td>0.000994</td>
<td>31.564578913</td>
<td>57.674</td>
<td>112.77</td>
<td>1</td>
<td>rs148500993</td>
</tr>
<tr>
<td>13:76754250:C:T</td>
<td>rs550007917</td>
<td>13</td>
<td>76754250</td>
<td>C</td>
<td>T</td>
<td>0.000994</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>rs148500993</td>
</tr>
</tbody>
</table>

uniqID: Unique ID of SNPs consists of chr:position:allele1:allele2 where alleles are alphabetically ordered.
rsID: rsID based on dbSNP build 146.
chr: Chromosome
pos: Position on hg19
effect allele: Effect/risk allele if it is provided in the input GWAS summary statistics file. If not, this is the alternative (minor) allele in 1000G.
non-effect allele: Non-effect/non-risk allele if it is provided in the input GWAS summary statistics file. If not, this is the reference (major) allele in 1000G.
MAF: Minor allele frequency computed based on 1000G.
gwasP: P-value provided in the input GWAS summary statistics file. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have 'NA' instead.
beta: Beta provided in the input GWAS summary statistics file if available. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have 'NA' instead.
SE: Standard error provided in the input GWAS summary statistics file if available. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have 'NA' instead.
r2: The maximum r2 of the SNP with one of the independent significant SNP (this doesn't have to be top lead SNPs in the genomic loci).
IndSigSNP: rsID of an independent significant SNP which has the maximum r2 of the SNP.
Appendix 2 All SNPs in LD with any of independent lead SNPs with r^2 greater or equal to the defined threshold (part 2)

<table>
<thead>
<tr>
<th>uniqID</th>
<th>genomic locus</th>
<th>nearest gene</th>
<th>dist</th>
<th>func</th>
<th>CADD</th>
<th>RDB</th>
<th>minChrState</th>
<th>common ChrState</th>
<th>posMapFilt</th>
<th>eqtlMapFilter</th>
<th>ciMapFilter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:50993178:C:T</td>
<td>1:50993178</td>
<td>FAF1</td>
<td>0</td>
<td>intronic</td>
<td>13.92</td>
<td>2b</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1:51128777:A:G</td>
<td>1:51128777</td>
<td>FAF1</td>
<td>0</td>
<td>intronic</td>
<td>2.789</td>
<td>6</td>
<td>5</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1:51142556:C:T</td>
<td>1:51142556</td>
<td>FAF1</td>
<td>0</td>
<td>intronic</td>
<td>1.043</td>
<td>6</td>
<td>5</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1:51161263:A:G</td>
<td>1:51161263</td>
<td>FAF1</td>
<td>0</td>
<td>intronic</td>
<td>3.298</td>
<td>6</td>
<td>5</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1:51264933:C:T</td>
<td>1:51264933</td>
<td>FAF1</td>
<td>0</td>
<td>intronic</td>
<td>9.294</td>
<td>5</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:28609064:A:C</td>
<td>4:28609064</td>
<td>RP11-123O22.1</td>
<td>7166</td>
<td>intergenic</td>
<td>0.04</td>
<td>NA</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:28641266:A:T</td>
<td>4:28641266</td>
<td>RP11-123O22.1</td>
<td>39368</td>
<td>intergenic</td>
<td>8.69</td>
<td>NA</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:28705788:C:G</td>
<td>4:28705788</td>
<td>RN7SL101P</td>
<td>7031</td>
<td>intergenic</td>
<td>2.049</td>
<td>NA</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:28936170:A:C</td>
<td>4:28936170</td>
<td>RP11-292B1.2</td>
<td>61975</td>
<td>intergenic</td>
<td>3.008</td>
<td>NA</td>
<td>9</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:28937541:C:T</td>
<td>4:28937541</td>
<td>RP11-292B1.2</td>
<td>60604</td>
<td>intergenic</td>
<td>0.681</td>
<td>NA</td>
<td>9</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:29047725:A:C</td>
<td>4:29047725</td>
<td>RP11-769N22.1</td>
<td>487</td>
<td>upstream</td>
<td>2.63</td>
<td>6</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:29094291:C:T</td>
<td>4:29094291</td>
<td>RP11-472K22.1</td>
<td>25638</td>
<td>intergenic</td>
<td>0.588</td>
<td>NA</td>
<td>9</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:29096073:C:G</td>
<td>4:29096073</td>
<td>RP11-472K22.1</td>
<td>23856</td>
<td>intergenic</td>
<td>1.162</td>
<td>NA</td>
<td>7</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:29436739:C:T</td>
<td>4:29436739</td>
<td>RP11-68D16.1</td>
<td>27104</td>
<td>intergenic</td>
<td>0.24</td>
<td>NA</td>
<td>9</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:29485633:C:T</td>
<td>4:29485633</td>
<td>RP11-68D16.2</td>
<td>16160</td>
<td>intergenic</td>
<td>0.987</td>
<td>NA</td>
<td>9</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:29807162:A:G</td>
<td>4:29807162</td>
<td>AC109351.1</td>
<td>55229</td>
<td>intergenic</td>
<td>0.337</td>
<td>NA</td>
<td>1</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4:30094647:C:G</td>
<td>4:30094647</td>
<td>RP11-174E22.2</td>
<td>84708</td>
<td>intergenic</td>
<td>0.289</td>
<td>NA</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13:74987745:C:T</td>
<td>13:74987745</td>
<td>AL355390.1</td>
<td>0</td>
<td>UTR3</td>
<td>0.052</td>
<td>7</td>
<td>5</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13:75032941:G:T</td>
<td>13:75032941</td>
<td>LINC00381</td>
<td>23644</td>
<td>intergenic</td>
<td>0.368</td>
<td>NA</td>
<td>9</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13:75185289:C:T</td>
<td>13:75185289</td>
<td>LINC00347</td>
<td>45706</td>
<td>intergenic</td>
<td>2.169</td>
<td>NA</td>
<td>9</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13:75220378:C:T</td>
<td>13:75220378</td>
<td>LINC00347</td>
<td>80795</td>
<td>intergenic</td>
<td>16.63</td>
<td>5</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>uniqID</td>
<td>genomic locus</td>
<td>nearest gene</td>
<td>dist</td>
<td>func</td>
<td>CADD</td>
<td>RDB</td>
<td>minChrState</td>
<td>commonChrState</td>
<td>posMapFilt</td>
<td>eqtlMapFilt</td>
<td>ciMapFilt</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------</td>
<td>--------------</td>
<td>------</td>
<td>------------</td>
<td>------</td>
<td>-----</td>
<td>-------------</td>
<td>----------------</td>
<td>------------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>13:75289942:A:G</td>
<td>3</td>
<td>RIOK3P1</td>
<td>113747</td>
<td>intergenic</td>
<td>4.54</td>
<td>NA</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13:75477615:G:T</td>
<td>3</td>
<td>RIOK3P1</td>
<td>73397</td>
<td>intergenic</td>
<td>8.447</td>
<td>NA</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13:75823871:C:G</td>
<td>3</td>
<td>LINC01078</td>
<td>744</td>
<td>upstream</td>
<td>2.942</td>
<td>6</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13:75873067:G:T</td>
<td>3</td>
<td>TBC1D4</td>
<td>0</td>
<td>intronic</td>
<td>0.197</td>
<td>NA</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13:76754250:C:T</td>
<td>3</td>
<td>RN7SL571P</td>
<td>44889</td>
<td>intergenic</td>
<td>0.633</td>
<td>NA</td>
<td>14</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**genomic locus:** Index of the genomic risk loci matching with "GenomicRiskLoci.txt".

**nearest gene:** The nearest Gene of the SNP based on ANNOVAR annotations.

Note that ANNOVAR annotates "consequence" function by prioritising the most deleterious annotation for SNPs which are locating a genomic region where multiple genes are overlapped.

Genes are encoded in symbol, if it is available, otherwise Ensembl ID.

Genes include all transcripts from Ensembl gene build 85 including non-protein coding genes and RNAs.

**dist:** Distance to the nearest gene. SNPs which are locating in the gene body or 1kb up- or down-stream of TSS or TES have 0.

**func:** Functional consequence of the SNP on the gene obtained from ANNOVAR. For exonic SNPs, detail annotation (e.g., non-synonymous, stop gain and so on) is available in ANNOVAR table (annov.txt).

**CADD:** CADD score which is computed based on 63 annotations. The higher score, the more deleterious the SNP is. 12.37 is the suggested threshold by Kicher et al (2014).

**RDB:** RegulomeDB score which is the categorical score (from 1a to 7). 1a is the highest score that the SNP has the most biological evidence to be regulatory element.

**minChrState:** The minimum 15-core chromatin state across 127 tissue/cell type.

**commonChrState:** The most common 15-core chromatin state across 127 tissue/cell types.

**posMapFilt:** Whether the SNP was used for positional mapping or not. 1 is used, otherwise 0. When positional mapping is not performed, all SNPs have 0.

**eqtlMapFilt:** Whether the SNP was used for eQTL mapping or not. 1 is used, otherwise 0. When eQTL mapping is not performed, all SNPs have 0.